GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 2002, 15:37:32 ; Search time 34.17 Seconds
(without alignments)
2080.399 Million cell updates/sec

Title: US-09-905-056-292
Perfect score: 3362
Sequence: 1 MLNKMTLHPQQIMIGPRFNR.....VHEPLLIRMNSKDNVQETQI 640

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3362	100.0	640	20	AAY08100	Human PRO331 prote
2	3362	100.0	640	20	AAW85722	Novel protein (Clo
3	3362	100.0	640	20	AAY13394	Amino acid sequenc
4	3362	100.0	640	21	AAB24407	Human PRO331 prote
5	3362	100.0	640	21	AAY70673	Human PRO331 prote
6	3362	100.0	640	22	AAU12355	Human PRO331 polyp
7	3362	100.0	640	22	AAU00826	Human immune respo
8	3362	100.0	640	22	AAB80262	Human PRO331 prote
9	3362	100.0	640	22	AAB65292	Human PRO331 prote
10	3362	100.0	640	22	AAB53089	Human angiogenesis
11	2131.5	63.4	713	22	AAE13006	Human leucine-rich
12	1865	55.5	653	21	AAB24073	Human PRO1111 prot
13	1865	55.5	653	21	AAY66694	Membrane-bound pro
14	1865	55.5	653	22	AAE09438	Human sbgPRO331a p
15	1865	55.5	653	22	AAU12390	Human PRO1111 poly
16	1865	55.5	653	22	AAB65217	Human PRO1111 (UNQ
17	1856	55.2	653	20	AAY28806	cc359_4 secreted p
18	1838	54.7	606	22	AAU18035	Human immunoglobul
19	1832	54.5	694	21	AAB23033	Human SLIT protein
20	1710.5	50.9	590	21	AAB23034	Human SLIT protein
21	1634	48.6	553	22	ABB10349	Human cDNA SEQ ID
22	1561.5	46.4	448	21	AAB23044	Human SLIT protein
23	1164	34.6	441	21	AAB43091	Human ORFX ORF2855
24	974	29.0	302	21	AAY87066	Human secreted pro
25	974	29.0	302	22	AAE06043	Human gene 3 encod
26	845.5	25.1	281	22	ABB10522	Human cDNA SEQ ID
27	845.5	25.1	281	22	AAU18084	Human immunoglobul
28	688	20.5	224	21	AAY87135	Human secreted pro
29	688	20.5	224	22	AAE06112	Human gene 3 encod
30	586.5	17.4	614	20	AAW84596	Amino acid sequenc
31	583.5	17.4	620	22	AAB74705	Human membrane ass
32	575	17.1	606	22	AAE06799	Human neuronal gui
33	574.5	17.1	620	20	AAY13357	Amino acid sequenc
34	574.5	17.1	620	22	AAU12333	Human PRO227 polyp
35	574.5	17.1	620	22	AAB80225	Human PRO227 prote
36	573.5	17.1	579	22	AAE06804	Mature human neuro
37	538.5	16.0	592	22	AAE09437	Human sbgTango79a
38	487	14.5	548	22	AAB31161	Amino acid sequenc
39	481.5	14.3	766	22	AAG67505	Amino acid sequenc
40	466	13.9	1021	22	ABB60162	Drosophila melanog
41	446	13.3	771	22	AAG67512	Amino acid sequenc
42	438.5	13.0	713	20	AAY13385	Amino acid sequenc
43	438.5	13.0	713	22	AAB80253	Human PRO293 prote
44	435.5	13.0	785	21	AAB12448	Human hh00149 prot
45	435.5	13.0	789	21	AAB09968	Human brain-specif

best post-art

ALIGNMENTS

RESULT 1
AAY08100
ID AAY08100 standard; Protein; 640 AA.
XX
AC AAY08100;
XX
DT 11-SEP-2000 (first entry)
XX
DE Human PRO331 protein.
XX
KW Inflammatory cell infiltration; immune response; T cell proliferation;
KW anti-inflammatory; anti-autoimmune; anti-diabetic; spondyloarthropathy;
KW T cell-mediated disease; spondyloarthropathy; sclerosis; renal disease;
KW inflammatory myopathy; hemolytic anemia; thrombocytopenia; thyroiditis;
KW diabetes mellitus; demyelinating polyneuropathy; Guillain-Barre syndrome;
KW multiple sclerosis; polyneuropathy; hepatitis; cirrhosis; enteropathy;
KW sclerosing cholangitis; inflammatory bowel disease; Whipple's disease;
KW skin disease; dermatitis; psoriasis; asthma; allergic rhinitis; tumor;
KW food hypersensitivity; urticaria; eosinophilic pneumonia; transplant;
KW idiopathic pulmonary fibrosis; graft rejection; PRO245; human;
KW PRO331.
XX
OS Homo sapiens.
XX
PN WO9914241-A2.
XX
PD 25-MAR-1999.
XX
PF 17-SEP-1998; 98WO-US19437.
XX
PR 17-SEP-1997; 97US-0059119.
PR 18-SEP-1997; 97US-0059263.
PR 28-OCT-1997; 97US-0063550.
PR 12-NOV-1997; 97US-0065186.

PR 21-NOV-1997; 97US-0066364.
PR 24-NOV-1997; 97US-0066770.
PR 04-JUN-1998; 98US-0088026.
XX
PA (GETH) GENENTECH INC.
XX
PI Fong S, Goddard A, Gurney AL, Tumas D, Wood WI;
XX
DR WPI; 1999-229499/19.
DR N-PSDB; AAX37738.
XX
PT Composition containing novel polypeptide PRO245, its agonist or
PT antagonist -
XX
PS Example 1; Fig 33; 177pp; English.
XX
CC This invention describes a novel composition containing (apart from a
CC carrier or excipient), a novel PRO245 polypeptide (I), its agonist or
CC antagonist, or their fragments, for modulating: (i) infiltration of
CC inflammatory cells into tissue; (ii) an immune response; or (iii) T cell
CC proliferation. The composition increases or decreases any of the effects
CC (i)-(iii). The products of the invention have anti-inflammatory,
CC anti-autoimmune and anti-diabetic activity. (I), and its (ant)agonists
CC and their fragments, are used to treat immune-related diseases,
CC particularly T cell-mediated diseases. The diseases treated include
CC systemic lupus erythematosus, rheumatoid arthritis, juvenile chronic
CC arthritis, spondyloarthropathies, systemic sclerosis (scleroderma),
CC idiopathic inflammatory myopathies (dermatomyositis, polymyositis),
CC Sjogren's syndrome, systemic vasculitis, sarcoidosis, autoimmune
CC hemolytic anemia (immune pancytopenia, paroxysmal nocturnal
CC hemoglobinuria), autoimmune thrombocytopenia (idiopathic thrombocytopenic
CC purpura immune-mediated thrombocytopenia), thyroiditis (Grave's disease,
CC Hashimoto's thyroiditis, juvenile lymphocytic thyroiditis, atrophic
CC thyroiditis), diabetes mellitus, immune-mediated renal disease
CC (glomerulonephritis, tubulointerstitial nephritis), multiple sclerosis,
CC idiopathic demyelinating polyneuropathy, Guillain-Barre syndrome, chronic
CC inflammatory demyelinating polyneuropathy, infectious hepatitis
CC (hepatitis A, B, C, D, E and other non-hepatotropic viruses), autoimmune
CC chronic active hepatitis, primary biliary cirrhosis, granulomatous
CC hepatitis, and sclerosing cholangitis, inflammatory bowel disease
CC (ulcerative colitis: Crohn's disease), gluten-sensitive enteropathy, and
CC Whipple's disease. Autoimmune or immune-mediated skin diseases including
CC bullous skin diseases, erythema multiforme,contact dermatitis, psoriasis,
CC asthma, allergic rhinitis, atopic dermatitis, food hypersensitivity,
CC urticaria, eosinophilic pneumonia, idiopathic pulmonary fibrosis,
CC hypersensitivity pneumonitis, and transplantation associated diseases
CC (graft rejection, and graft-versus-host-disease). (I), its (ant)agonists
CC or fragment can also be used as an adjuvant in treatment of tumors.
CC Antibodies against (I) can also be used for diagnosing such diseases.
CC This sequence represents the human PRO331 protein which is described in
CC the method of the invention.
XX
SQ Sequence 640 AA;

Query Match 100.0%; Score 3362; DB 20; Length 640;
Best Local Similarity 100.0%; Pred. No. 2.3e-234;
Matches 640; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MLNKMTLHPQQIMIGPRFNRALFDPLLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mlnkmtlhpqqimigprfnralfdpllvvllalqllvvaglvraqtcpsvcscsnqfskv 60

Qy     61 ICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAF 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 icvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhleilqlsrnhirtieigaf 120

Qy    121 NGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 nglanlntlelfdnrlttipngafvylsklkelwlrnnpiesipsyafnripslrrldlg 180

Qy    181 ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQ 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 elkrlsyisegafeglsnlrylnlamcnlreipnltplikldeldlsgnhlsairpgsfq 240

Qy    241 GLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHH 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 glmhlqklwmiqsqiqviernafdnlqslveinlahnnltllphdlftplhhlerihlhh 300

Qy    301 NPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPP 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 npwncncdilwlswwikdmapsntaccarcntppnlkgryigeldqnyftcyapvivepp 360

Qy    361 ADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQD 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 adlnvtegmaaelkcrastsltsvswitpngtvmthgaykvriavlsdgtlnftnvtvqd 420

Qy    421 TGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTP 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 tgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmepsqdearttdnnvgptp 480

Qy    481 VVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGCFVAITLMA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 vvdwettnvttsltpqstrstektftipvtdinsgipgidevmkttkiiigcfvaitlma 540

Qy    541 AVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPMPAIEHEHLNHYNS 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 avmlvifykmrkqhhrqnhhaptrtveiinvddeitgdtpmeshlpmpaiehehlnhyns 600

Qy    601 YKSPFNHTTTVNTINSIHSSVHEPLLIRMNSKDNVQETQI 640
          ||||||||||||||||||||||||||||||||||||||||
Db    601 ykspfnhtttvntinsihssvhepllirmnskdnvqetqi 640
```

RESULT 2
AAW85722
ID AAW85722 standard; Protein; 640 AA.
XX
AC AAW85722;
XX
DT 27-SEP-1999 (first entry)
XX
DE Novel protein (Clone AS209_1).
XX
KW Polynucleotide; protein; nutrition; cytokine; cell proliferation;
KW cell differentiation; immunostimulation; immunosuppression;
KW haematopoiesis regulation; tissue growth; activin; inhibin;
KW chemotaxis; chemokinesis; haemostasis; thrombolysis; receptor;
KW ligand; anti-inflammatory; tumour suppression; gene therapy.
XX
OS Homo sapiens.
XX
PN WO9920644-A1.
XX
PD 29-APR-1999.
XX
PF 16-OCT-1998; 98WO-US22034.
XX
PR 18-OCT-1997; 97US-0955557.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Agostino MJ, Bowman MR, Evans C, Jacobs K, Lavallie ER;
PI McCoy JM, Merberg D, Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1999-288272/24.
DR N-PSDB; AAX08687.
XX
PT New polynucleotides encoding secreted human proteins
XX
PS Claim 26; Page 109-111; 136pp; English.
XX
CC The new human secreted proteins are encoded by polynucleotides
CC obtained from human placenta, adult testes, fetal kidney, fetal
CC brain, adult brain, adult brain and adult blood cDNA libraries.
CC The polynucleotides and proteins are predicted to have biological

```
CC   activities which would make them suitable for treating, preventing or
CC   ameliorating medical conditions in humans and animals. Suggested
CC   activities include nutritional activity, cytokine and cell
CC   proliferation/differentiation activity, immune stimulating (e.g. as
CC   vaccines) or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, and tumour inhibition
CC   activity.  The polynucleotides are also stated to be useful for gene
CC   therapy. The sequences identified by a secretory leader
CC   sequence motif in the polynucleotide and it is thought that the
CC   encoded proteins have biological activity by virtue of their secreted
CC   nature. This polypeptide was encoded by a clone designated AS209_1
CC   (See AAX08687).
XX
SQ   Sequence   640 AA;


  Query Match             100.0%;  Score 3362;  DB 20;  Length 640;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-234;
  Matches  640;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MLNKMTLHPQQIMIGPRFNRALFDPLLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mlnkmtlhpqqimigprfnralfdpllvvllalqllvvaglvraqtcpsvcscsnqfskv 60

Qy     61 ICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAF 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 icvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhleilqlsrnhirtieigaf 120

Qy    121 NGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 nglanlntlelfdnrlttipngafvylsklkelwlrnnpiesipsyafnripslrrldlg 180

Qy    181 ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQ 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 elkrlsyisegafeglsnlrylnlamcnlreipnltplikldeldlsgnhlsairpgsfq 240

Qy    241 GLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHH 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 glmhlqklwmiqsqiqviernafdnlqslveinlahnnltllphdlftplhhlerihlhh 300

Qy    301 NPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPP 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 npwncncdilwlswwikdmapsntaccarcntppnlkgryigeldqnyftcyapvivepp 360

Qy    361 ADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQD 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 adlnvtegmaaelkcrastsltsvswitpngtvmthgaykvriavlsdgtlnftnvtvqd 420

Qy    421 TGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTP 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 tgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmepsqdearttdnnvgptp 480

Qy    481 VVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGCFVAITLMA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 vvdwettnvttsltpqstrstektftipvtdinsgipgidevmkttkiiigcfvaitlma 540

Qy    541 AVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPMPAIEHEHLNHYNS 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 avmlvifykmrkqhhrqnhhaptrtveiinvddeitgdtpmeshlpmpaiehehlnhyns 600

Qy    601 YKSPFNHTTTVNTINSIHSSVHEPLLIRMNSKDNVQETQI 640
          ||||||||||||||||||||||||||||||||||||||||
Db    601 ykspfnhtttvntinsihssvhepllirmnskdnvqetqi 640


RESULT   3
AAY13394
ID   AAY13394 standard; Protein; 640 AA.
XX
AC   AAY13394;
XX
DT   25-JUN-1999  (first entry)
XX
DE   Amino acid sequence of protein PRO331.
XX
KW   Secreted protein; transmembrane protein; human; enterocolitis;
KW   Zollinger-Ellison syndrome; gastrointestinal ulceration;
KW   congenital microvillus atrophy; skin disease; cell growth;
KW   abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
KW   Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
KW   fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;
KW   anti-thrombotic; wound healing; tissue repair.
XX
OS   Homo sapiens.
XX
PN   WO9914328-A2.
XX
PD   25-MAR-1999.
XX
PF   16-SEP-1998;   98WO-US19330.
XX
PR   25-NOV-1997;   97US-0066840.
PR   17-SEP-1997;   97US-0059113.
PR   17-SEP-1997;   97US-0059115.
PR   17-SEP-1997;   97US-0059117.
PR   17-SEP-1997;   97US-0059119.
PR   17-SEP-1997;   97US-0059121.
PR   17-SEP-1997;   97US-0059122.
PR   17-SEP-1997;   97US-0059184.
PR   18-SEP-1997;   97US-0059263.
PR   18-SEP-1997;   97US-0059266.
PR   15-OCT-1997;   97US-0062125.
PR   17-OCT-1997;   97US-0062285.
PR   17-OCT-1997;   97US-0062287.
PR   21-OCT-1997;   97US-0063486.
PR   24-OCT-1997;   97US-0062814.
PR   24-OCT-1997;   97US-0062816.
PR   24-OCT-1997;   97US-0063045.
PR   24-OCT-1997;   97US-0063120.
PR   24-OCT-1997;   97US-0063121.
PR   24-OCT-1997;   97US-0063127.
PR   24-OCT-1997;   97US-0063128.
PR   27-OCT-1997;   97US-0063329.
PR   27-OCT-1997;   97US-0063327.
PR   28-OCT-1997;   97US-0063541.
PR   28-OCT-1997;   97US-0063542.
PR   28-OCT-1997;   97US-0063544.
PR   28-OCT-1997;   97US-0063549.
PR   28-OCT-1997;   97US-0063550.
PR   28-OCT-1997;   97US-0063564.
PR   29-OCT-1997;   97US-0063435.
PR   29-OCT-1997;   97US-0063704.
PR   29-OCT-1997;   97US-0063732.
PR   29-OCT-1997;   97US-0063738.
PR   29-OCT-1997;   97US-0063734.
PR   29-OCT-1997;   97US-0064215.
PR   29-OCT-1997;   97US-0063735.
PR   31-OCT-1997;   97US-0063870.
PR   31-OCT-1997;   97US-0064103.
PR   03-NOV-1997;   97US-0064248.
PR   07-NOV-1997;   97US-0064809.
PR   12-NOV-1997;   97US-0065186.
PR   17-NOV-1997;   97US-0065846.
PR   18-NOV-1997;   97US-0065693.
PR   21-NOV-1997;   97US-0066120.
PR   21-NOV-1997;   97US-0066364.
PR   24-NOV-1997;   97US-0066772.
PR   24-NOV-1997;   97US-0066466.
PR   24-NOV-1997;   97US-0066770.
PR   24-NOV-1997;   97US-0066511.
PR   24-NOV-1997;   97US-0066453.
XX
PA   (GETH ) GENENTECH INC.
```

```
XX
PI   Chen J,  Goddard A,  Gurney AL,  Pennica D,  Wood WI,  Yuan J;
XX
DR   WPI; 1999-229533/19.
DR   N-PSDB; AAX52265.
XX
PT   New isolated human genes and polypeptides used in, e.g. treatment of
PT   gastrointestinal ulceration
XX
PS   Claim 12; Fig 104; 320pp; English.
XX
CC   AAY13344-403 represent secreted and transmembrane human proteins.
CC   The cDNA sequences are obtained from cDNA libraries, prepared from
CC   fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
CC   The encoded polypeptides have specific uses based on their homology to
CC   known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
CC   associated with the preservation and maintenance of gastrointestinal
CC   mucosa and the repair of acute and chronic mucosal lesions
CC   (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
CC   ulceration and congenital microvillus atrophy), skin diseases associated
CC   with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
CC   cancers such as lung squamous cell carcinoma of the vulva and gliomas),
CC   potent effects on cell growth and development, diseases related to
CC   growth or survival of nerve cells including Parkinson's disease,
CC   Alzheimer's disease, ALS, neuropathies or cancer. PRO265 can be used as
CC   for fibromodulin, e.g. for reducing dermal scarring.  PRO264 can be used
CC   as a target for anti-tumor drugs. PRO533 may be used in the treatment
CC   of Usher Syndrome or Atrophia areata; PRO269 can be used as an
CC   anti-thrombotic agent; PRO287 polypeptides and portions may have
CC   therapeutic applications in wound healing and tissue repair; PRO317 can
CC   be used for treating problems of the kidney, uterus, endometrium, blood
CC   vessels, or related tissue, e.g. in the heart of genital tract.
XX
SQ   Sequence   640 AA;


  Query Match           100.0%;  Score 3362;  DB 20;  Length 640;
  Best Local Similarity  100.0%;  Pred. No. 2.3e-234;
  Matches  640;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MLNKMTLHPQQIMIGPRFNRALFDPLLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mlnkmtlhpqqimigprfnralfdpllvvllalqllvvaglvraqtcpsvcscsnqfskv 60

Qy     61 ICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAF 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 icvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhleilqlsrnhirtieigaf 120

Qy    121 NGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 nglanlntlelfdnrlttipngafvylsklkelwlrnnpiesipsyafnripslrrldlg 180

Qy    181 ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQ 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 elkrlsyisegafeglsnlrylnlamcnlreipnltplikldeldlsgnhlsairpgsfq 240

Qy    241 GLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHH 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 glmhlqklwmiqsqiqviernafdnlqslveinlahnnltllphdlftplhhlerihlhh 300

Qy    301 NPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPP 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 npwncncdilwlswwikdmapsntaccarcntppnlkgryigeldqnyftcyapvivepp 360

Qy    361 ADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQD 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 adlnvtegmaaelkcrastsltsvswitpngtvmthgaykvriavlsdgtlnftnvtvqd 420

Qy    421 TGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTP 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 tgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmepsqdearttdnnvgptp 480

Qy    481 VVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGCFVAITLMA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 vvdwettnvttsltpqstrstektftipvtdinsgipgidevmkttkiiigcfvaitlma 540

Qy    541 AVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPMPAIEHEHLNHYNS 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 avmlvifykmrkqhhrqnhhaptrtveiinvddeitgdtpmeshlpmpaiehehlnhyns 600

Qy    601 YKSPFNHTTTVNTINSIHSSVHEPLLIRMNSKDNVQETQI 640
          ||||||||||||||||||||||||||||||||||||||||
Db    601 ykspfnhtttvntinsihssvhepllirmnskdnvqetqi 640


RESULT   4
AAB24407
ID   AAB24407 standard; Protein; 640 AA.
XX
AC   AAB24407;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Human PRO331 protein sequence SEQ ID NO:107.
XX
KW   Human; PRO; promotion; inhibition; angiogenesis; cardiovascularisation;
KW   diagnosis; trauma; wound; cancer; atherosclerosis; cardiac hypertrophy;
KW   angiogenic; proliferative; cardiant; cardiovascular; antiatherosclerotic;
KW   cytostatic; gene therapy; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200032221-A2.
XX
PD   08-JUN-2000.
XX
PF   30-NOV-1999;   99WO-US28313.
XX
PR   01-DEC-1998;   98WO-US25108.
PR   16-DEC-1998;   98US-0112850.
PR   12-JAN-1999;   99US-0115554.
PR   08-MAR-1999;   99WO-US05028.
PR   12-MAR-1999;   99US-0123957.
PR   28-APR-1999;   99US-0131445.
PR   14-MAY-1999;   99US-0134287.
PR   02-JUN-1999;   99WO-US12252.
PR   23-JUN-1999;   99US-0141037.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   01-SEP-1999;   99WO-US20111.
PR   08-SEP-1999;   99WO-US20594.
PR   13-SEP-1999;   99WO-US20944.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   05-OCT-1999;   99WO-US23089.
PR   29-OCT-1999;   99US-0162506.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Ferrara N,  Gerber H,  Hillan KJ,  Goddard A;
PI   Godowski PJ,  Gurney AL,  Klein RD,  Kuo SS,  Paoni NF,  Smith V;
PI   Watanabe CK,  Williams PM,  Wood WI;
XX
DR   WPI; 2000-412154/35.
DR   N-PSDB; AAA77596.
XX
PT   Nucleic acids encoding PRO polypeptides useful for preventing,
PT   diagnosing and treating diagnosing a cardiovascular, endothelial or
PT   angiogenic disorders in mammals -
XX
PS   Claim 72; Fig 40; 315pp; English.
XX
CC   The present invention describes nucleic acids encoding PRO polypeptides
CC   useful for preventing, diagnosing and treating diagnosing a
CC   cardiovascular, endothelial or angiogenic disorder in mammals by
CC   modulating cell proliferation, angiogenesis and cardiovascularisation,
```

```
CC   and for identifying agonists and antagonists of these processes. The
CC   nucleic acids and the proteins they encode may be used in the
CC   prevention, treatment and diagnosis of diseases associated with
CC   inappropriate PRO expression such as cardiovascular, endothelial or
CC   angiogenic disorders in mammals (e.g. atherosclerosis, cancers and
CC   cardiac hypertrophy). For example, the nucleic acids (NCs) and vectors
CC   containing them and the PRO polypeptide may be used to treat disorders
CC   associated with decreased PRO expression. AAA77510 to AAA77721 and
CC   AAB24388 to AAB24435 represent nucleotide and protein sequences used in
CC   the exemplification of the present invention.
XX
SQ   Sequence   640 AA;


  Query Match           100.0%;  Score 3362;  DB 21;  Length 640;
  Best Local Similarity  100.0%;  Pred. No. 2.3e-234;
  Matches  640;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MLNKMTLHPQQIMIGPRFNRALFDPLLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mlnkmtlhpqqimigprfnralfdpllvvllalqllvvaglvraqtcpsvcscsnqfskv 60

Qy     61 ICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAF 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 icvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhleilqlsrnhirtieigaf 120

Qy    121 NGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 nglanlntlelfdnrlttipngafvylsklkelwlrnnpiesipsyafnripslrrldlg 180

Qy    181 ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQ 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 elkrlsyisegafeglsnlrylnlamcnlreipnltplikldeldlsgnhlsairpgsfq 240

Qy    241 GLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHH 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 glmhlqklwmiqsqiqviernafdnlqslveinlahnnltllphdlftplhhlerihlhh 300

Qy    301 NPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPP 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 npwncncdilwlswwikdmapsntaccarcntppnlkgryigeldqnyftcyapvivepp 360

Qy    361 ADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQD 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 adlnvtegmaaelkcrastsltsvswitpngtvmthgaykvriavlsdgtlnftnvtvqd 420

Qy    421 TGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTP 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 tgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmepsqdearttdnnvgptp 480

Qy    481 VVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGCFVAITLMA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 vvdwettnvttsltpqstrstektftipvtdinsgipgidevmkttkiiigcfvaitlma 540

Qy    541 AVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPMPAIEHEHLNHYNS 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 avmlvifykmrkqhhrqnhhaptrtveiinvddeitgdtpmeshlpmpaiehehlnhyns 600

Qy    601 YKSPFNHTTTVNTINSIHSSVHEPLLIRMNSKDNVQETQI 640
          ||||||||||||||||||||||||||||||||||||||||
Db    601 ykspfnhtttvntinsihssvhepllirmnskdnvqetqi 640


RESULT   5
AAY70673
ID   AAY70673 standard; Protein; 640 AA.
XX
AC   AAY70673;
XX
DT   18-JUL-2000  (first entry)
XX
DE   Human PRO331 protein.
XX
KW   PRO331; UNQ292; dermatological; immunosuppressive; antiinflammatory;
KW   immunostimulant; antiasthmatic; antirheumatic; antiarthritic; virucide;
KW   antiallergic; haemostatic; hepatotropic; antidiabetic; antianaemic;
KW   nephrotropic; neuroprotective; anticoagulant; immunological disorder;
KW   lung; pneumonia; skin; psoriasis; kidney; glomerulonephritis; arthritis;
KW   spondyloarthropathy; SLE; systemic lupus erythematosis; scleroderma;
KW   idiopathic inflammatory myopathy; anaemia; thrombocytopenia; diabetes;
KW   thyroiditis; Grave's disease; demyelinating disease; multiple sclerosis;
KW   Crohn's disease; hepatobiliary disease; hepatitis; asthma; human;
KW   graft-versus-host-disease.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   40..46
FT                   /note= "N-myristoylation site"
FT   Modified-site   73..79
FT                   /note= "N-myristoylation site"
FT   Modified-site   118..124
FT                   /note= "N-myristoylation site"
FT   Modified-site   183..187
FT                   /note= "cAMP and cGMP-dependent protein kinase
FT                   phosphorylation site"
FT   Modified-site   191..197
FT                   /note= "N-myristoylation site"
FT   Modified-site   228..234
FT                   /note= "N-myristoylation site"
FT   Modified-site   237..243
FT                   /note= "N-myristoylation site"
FT   Modified-site   268..272
FT                   /note= "Casein Kinase II phosphorylation site"
FT   Modified-site   278..282
FT                   /note= "N-glycosylation site"
FT   Modified-site   364..368
FT                   /note= "N-glycosylation site"
FT   Modified-site   390..394
FT                   /note= "N-glycosylation site"
FT   Modified-site   391..397
FT                   /note= "N-myristoylation site"
FT   Modified-site   412..416
FT                   /note= "N-glycosylation site"
FT   Modified-site   415..419
FT                   /note= "N-glycosylation site"
FT   Modified-site   417..421
FT                   /note= "Casein Kinase II phosphorylation site"
FT   Modified-site   422..428
FT                   /note= "N-myristoylation site"
FT   Modified-site   433..439
FT                   /note= "N-myristoylation site"
FT   Modified-site   434..438
FT                   /note= "N-glycosylation site"
FT   Modified-site   442..446
FT                   /note= "N-glycosylation site"
FT   Modified-site   465..469
FT                   /note= "Casein Kinase II phosphorylation site"
FT   Modified-site   488..492
FT                   /note= "N-glycosylation site"
FT   Modified-site   531..537
FT                   /note= "N-myristoylation site"
FT   Modified-site   579..583
FT                   /note= "Casein Kinase II phosphorylation site"
FT   Modified-site   606..610
FT                   /note= "N-glycosylation site"
FT   Modified-site   620..624
FT                   /note= "Casein Kinase II phosphorylation site"
XX
PN   WO200015797-A2.
XX
PD   23-MAR-2000.
XX
PF   15-SEP-1999;   99WO-US21547.
XX
PR   17-SEP-1998;   98US-0100858.
```

```
PR   17-SEP-1998;   98WO-US19437.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Fong S,  Goddard A,  Gurney AL,  Tumas D,  Wood WI;
XX
DR   WPI; 2000-271435/23.
DR   N-PSDB; AAZ52207.
XX
PT   Composition for treatment and diagnosis of immune related diseases e.g.
PT   Grave's disease comprises a PRO245, PRO217, PRO301, PRO266, PRO335,
PT   PRO331 or PRO326 polypeptide or its agonists or antagonists (preferably
PT   antibodies) -
XX
PS   Example 1; Fig 14; 201pp; English.
XX
CC   The present sequence is the human protein PRO331, encoded by UNQ292 cDNA,
CC   designated as clone DNA40981. It is isolated from human foetal brain
CC   tissue. Portions of PRO331 has homology to the LIG-1 protein.
CC   It enhances or suppresses the infiltration of inflammatory cells into
CC   tissues, proliferation of T-lymphocytes and modulates the immune
CC   response. This sequence is useful for treatment of immune related
CC   disorders, like SLE, rheumatoid/juvenile arthritis, spondyloarthropathy,
CC   systemic sclerosis (scleroderma), idiopathic inflammatory myopathies such
CC   as dermatomyositis, Sjogren's syndrome, systemic vasculitis, sarcoidosis,
CC   autoimmune haemolytic anaemia, thrombocytopenia, thyroiditis e.g. Grave's
CC   disease, diabetes mellitus, immune-mediated renal disease e.g.
CC   glomerulonephritis, demyelinating diseases such as multiple sclerosis and
CC   Guillain-Barre syndrome, hepatobiliary diseases like hepatitis and
CC   primary biliary cirrhosis, inflammatory and fibrotic lung diseases such
CC   as inflammatory bowel disease (e.g. Crohn's disease), autoimmune or
CC   immune-mediated skin diseases such as psoriasis, allergies like asthma,
CC   immunological diseases of the lungs such as eosinophilic pneumonia and
CC   transplantation associated diseases such as graft-versus-host-disease.
XX
SQ   Sequence   640 AA;

  Query Match           100.0%;  Score 3362;  DB 21;  Length 640;
  Best Local Similarity  100.0%;  Pred. No. 2.3e-234;
  Matches  640;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MLNKMTLHPQQIMIGPRFNRALFDPLLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mlnkmtlhpqqimigprfnralfdpllvvllalqllvvaglvraqtcpsvcscsnqfskv 60

Qy      61 ICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAF 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 icvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhleilqlsrnhirtieigaf 120

Qy     121 NGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 nglanlntlelfdnrlttipngafvylsklkelwlrnnpiesipsyafnripslrrldlg 180

Qy     181 ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQ 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 elkrlsyisegafeglsnlrylnlamcnlreipnltplikldeldlsgnhlsairpgsfq 240

Qy     241 GLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHH 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 glmhlqklwmiqsqiqviernafdnlqslveinlahnnltllphdlftplhhlerihlhh 300

Qy     301 NPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPP 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 npwncncdilwlswwikdmapsntaccarcntppnlkgryigeldqnyftcyapvivepp 360

Qy     361 ADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQD 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 adlnvtegmaaelkcrastsltsvswitpngtvmthgaykvriavlsdgtlnftnvtvqd 420

Qy     421 TGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTP 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 tgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmepsqdearttdnnvgptp 480

Qy     481 VVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGCFVAITLMA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 vvdwettnvttsltpqstrstektftipvtdinsgipgidevmkttkiiigcfvaitlma 540

Qy     541 AVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPMPAIEHEHLNHYNS 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 avmlvifykmrkqhhrqnhhaptrtveiinvddeitgdtpmeshlpmpaiehehlnhyns 600

Qy     601 YKSPFNHTTTVNTINSIHSSVHEPLLIRMNSKDNVQETQI 640
          ||||||||||||||||||||||||||||||||||||||||
Db     601 ykspfnhtttvntinsihssvhepllirmnskdnvqetqi 640


RESULT   6
AAU12355
ID   AAU12355 standard; Protein; 640 AA.
XX
AC   AAU12355;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human PRO331 polypeptide sequence.
XX
KW   Human secretory and transmembrane; PRO; mammalian; cancer; lung;
KW   breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW   cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW   adipocyte; A-peptide; factor VIIA; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200140466-A2.
XX
PD   07-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-US32678.
XX
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28564.
PR   02-DEC-1999;   99WO-US28565.
PR   09-DEC-1999;   99US-0170262.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   20-DEC-1999;   99WO-US30999.
PR   30-DEC-1999;   99WO-US31243.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   01-MAR-2000; 2000WO-US05601.
PR   20-MAR-2000; 2000WO-US07377.
PR   21-MAR-2000; 2000WO-US07532.
PR   30-MAR-2000; 2000WO-US08439.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   10-NOV-2000; 2000WO-US30873.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Beresini M,  Deforge L,  Desnoyers L,  Filvaroff E,  Gao W;
PI   Gerritsen ME,  Goddard A,  Godowski PJ,  Gurney AL,  Sherwood S;
PI   Smith V,  Stewart TA,  Tumas D,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2001-408281/43.
DR   N-PSDB; AAS21427.
```

```
XX
PT   Isolated, secretory and transmembrane PRO polypeptide used to detect
PT   other PRO polypeptides, link bioactive molecules to cells expressing
PT   PRO polypeptides, and detect the presence of mammalian tumours e.g.
PT   lung, breast, prostate, cervical  -
XX
PS   Claim 12; Fig 368; 813pp; English.
XX
CC   AAU12172-AAU12446 represent novel human secretory and transmembrane
CC   PRO polypeptides. The PRO polypeptides are useful to detect other
CC   PRO polypeptides, to link bioactive molecules to cells expressing
CC   PRO polypeptides, to modulate biological activities of cells expressing
CC   PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC   breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC   polypeptide expression in a cell sample to that in a control sample.
CC   Some of the 275 sequences are also useful to stimulate the release of
CC   tumour necrosis factor-alpha (TNF-alpha) from human blood, the
CC   proliferation or differentiation of chondrocytes, the proliferation or
CC   gene expression in pericyte cells, the release of proteoglycans from
CC   cartilage, the proliferation of inner ear utricular supporting cells or
CC   of T-lymphocytes, the release of a cytokine from peripheral blood
CC   monocytes (PBMCs), or the proliferation of endothelial cells. Some of
CC   the PRO polypeptides may modulate glucose or free fatty acid uptake by
CC   skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
CC   to factor VIIA. The PRO polypeptides can be used in assays to identify
CC   molecules involved in binding interactions. The polynucleotides encoding
CC   PRO polypeptides can be used to generate probes, antisense RNA/DNA,
CC   transgenic or knock out animals and can be used in gene therapy.
XX
SQ   Sequence   640 AA;


  Query Match           100.0%;  Score 3362;  DB 22;  Length 640;
  Best Local Similarity  100.0%;  Pred. No. 2.3e-234;
  Matches  640;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MLNKMTLHPQQIMIGPRFNRALFDPLLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mlnkmtlhpqqimigprfnralfdpllvvllalqllvvaglvraqtcpsvcscsnqfskv 60

Qy      61 ICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAF 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 icvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhleilqlsrnhirtieigaf 120

Qy     121 NGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 nglanlntlelfdnrlttipngafvylsklkelwlrnnpiesipsyafnripslrrldlg 180

Qy     181 ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQ 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 elkrlsyisegafeglsnlrylnlamcnlreipnltplikldeldlsgnhlsairpgsfq 240

Qy     241 GLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHH 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 glmhlqklwmiqsqiqviernafdnlqslveinlahnnltllphdlftplhhlerihlhh 300

Qy     301 NPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPP 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 npwncncdilwlswwikdmapsntaccarcntppnlkgryigeldqnyftcyapvivepp 360

Qy     361 ADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQD 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 adlnvtegmaaelkcrastsltsvswitpngtvmthgaykvriavlsdgtlnftnvtvqd 420

Qy     421 TGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTP 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 tgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmepsqdearttdnnvgptp 480

Qy     481 VVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGCFVAITLMA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 vvdwettnvttsltpqstrstektftipvtdinsgipgidevmkttkiiigcfvaitlma 540

Qy     541 AVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPMPAIEHEHLNHYNS 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 avmlvifykmrkqhhrqnhhaptrtveiinvddeitgdtpmeshlpmpaiehehlnhyns 600

Qy     601 YKSPFNHTTTVNTINSIHSSVHEPLLIRMNSKDNVQETQI 640
           ||||||||||||||||||||||||||||||||||||||||
Db     601 ykspfnhtttvntinsihssvhepllirmnskdnvqetqi 640


RESULT   7
AAU00826
ID   AAU00826 standard; Protein; 640 AA.
XX
AC   AAU00826;
XX
DT   04-JUL-2001  (first entry)
XX
DE   Human immune response protein PRO331 (UNQ292).
XX
KW   Human; PRO331; UNQ292; immune response; osteoarthritis;
KW   systemic lupus erythematosus; rheumatoid arthritis; systemic sclerosis;
KW   juvenile chronic arthritis; spondyloarthropathy; Sjogren's syndrome;
KW   idiopathic inflammatory; myopathy; polymyositis; systemic vasculitis;
KW   sarcoidosis; autoimmune haemolytic anaemia; immune pancytopaenia;
KW   autoimmune thrombocytopaenia; idiopathic thrombocytopaenic purpura;
KW   thyroiditis; Grave's disease; Hashimoto's thyroiditis;
KW   diabetes mellitus; glomerulonephritis; demyelinating disease;
KW   multiple sclerosis; Guillain-Barre syndrome; hepatobiliary disease;
KW   chronic inflammatory demyelinating polyneuropathy; infectious hepatitis;
KW   auto immune chronic active hepatitis; primary biliary cirrhosis;
KW   granulomatous hepatitis; sclerosing cholangitis; ulcerative colitis;
KW   inflammatory bowel disease; Crohn's disease; Whipple's disease;
KW   erythaemia multiforme; psoriasis; asthma; allergic rhinitis; urticaria;
KW   food hypersensitivity; eosinophilic pneumonia; graft rejection;
KW   idiopathic pulmonary fibrosis; graft-versus-host-disease; immunogen;
KW   antibody.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..44
FT                   /label= Signal_peptide
FT   Modified-site   40..46
FT                   /note= "Glycine is N-myristoylated"
FT   Protein         45..640
FT                   /label= Mature_PRO331
FT   Modified-site   73..79
FT                   /note= "Glycine is N-myristoylated"
FT   Modified-site   118..124
FT                   /note= "Glycine at 118 is N-myristoylated"
FT   Region          183..187
FT                   /label= Phosphorylation_site
FT                   /note= "cAMP/cGMP dependent protein kinase
FT                   phosphorylation site"
FT   Modified-site   191..197
FT                   /note= "Glycine at 191 is N-myristoylated"
FT   Modified-site   228..234
FT                   /note= "Glycine is N-myristoylated"
FT   Modified-site   237..243
FT                   /note= "Glycine is N-myristoylated"
FT   Modified-site   278..282
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   364..368
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   390..394
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   391..397
FT                   /note= "Glycine at 391 is N-myristoylated"
FT   Modified-site   412..416
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   415..419
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   422..428
FT                   /note= "Glycine is N-myristoylated"
```

```
FT   Modified-site   433..439
FT                   /note= "Glycine is N-myristoylated"
FT   Modified-site   434..438
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   442..446
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   488..492
FT                   /note= "Asn is N-glycosylated"
FT   Domain          528..543
FT                   /label= Transmembrane_domain
FT   Modified-site   531..537
FT                   /note= "Glycine is N-myristoylated"
FT   Modified-site   606..610
FT                   /note= "Asn is N-glycosylated"
XX
PN   WO200119991-A1.
XX
PD   22-MAR-2001.
XX
PF   20-MAR-2000; 2000WO-US07377.
XX
PR   15-SEP-1999;   99WO-US21547.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Fong S,  Goddard A,  Gurney AL,  Hillan KJ,  Tumas D,  Wood WI;
XX
DR   WPI; 2001-226823/23.
DR   N-PSDB; AAS00162.
XX
PT   Composition for diagnosing and treating immune related diseases, e.g.,
PT   rheumatoid arthritis and diabetes mellitus, comprises a PRO
PT   polypeptide, agonist, antagonist or fragment -
XX
PS   Claim 31; Fig 12; 138pp; English.
XX
CC   The sequence represents Human PRO331 (UNQ292), a protein involved in
CC   the immune response. PRO polypeptides, and (ant)agonists to them, are
CC   used in compositions for modulating infiltration of inflammatory cells
CC   into a tissue, modulating an immune response and modulating proliferation
CC   of T-lymphocytes in response to an antigen. Immune related diseases can
CC   be treated with the compositions, such as, systemic lupus erythematosus,
CC   rheumatoid arthritis, osteoarthritis, juvenile chronic arthritis,
CC   spondyloarthopathies, systemic sclerosis, idiopathic inflammatory
CC   myopathies (e.g. polymyositis), Sjogren's syndrome, systemic vasculitis,
CC   sarcoidosis, autoimmune haemolytic anaemia (e.g. immune pancytopaenia),
CC   autoimmune thrombocytopaenia (e.g. idiopathic thrombocytopaenic purpura),
CC   thyroiditis (e.g. Grave's disease, Hashimoto's thyroiditis), diabetes
CC   mellitus, immune-mediated renal disease (e.g. glomerulonephritis),
CC   demyelinating diseases of the central and peripheral nervous systems e.g.
CC   multiple sclerosis or Guillain-Barre syndrome, and chronic inflammatory
CC   demyelinating polyneuropathy, hepatobiliary diseases such as infectious
CC   hepatitis (hepatitis A, B, C, D, E and other non-hepatotropic viruses),
CC   auto immune chronic active hepatitis, primary biliary cirrhosis,
CC   granulomatous hepatitis, and sclerosing cholangitis, inflammatory bowel
CC   disease (ulcerative colitis, Crohn's disease and Whipple's disease),
CC   autoimmune or immune-mediated skin diseases (e.g. erythaemia
CC   multiforme and psoriasis), asthma, allergic rhinitis, urticaria,
CC   food hypersensitivity, immunologic diseases of the lung such as
CC   eosinophilic pneumonias, idiopathic pulmonary fibrosis, transplantation
CC   associated diseases including graft-versus-host-disease and graft
CC   rejection. PRO polypeptides can be used to diagnose immune related
CC   diseases, to identify inhibitors, and to stimulate the proliferation of
CC   T lymphocytes.  Anti-PRO antibodies can be used to detect PRO and
CC   in diagnosis. PRO polypeptides, antibodies and (ant)agonists can be used
CC   in rational drug design.
XX
SQ   Sequence   640 AA;

  Query Match             100.0%;  Score 3362;  DB 22;  Length 640;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-234;
  Matches  640;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MLNKMTLHPQQIMIGPRFNRALFDPLLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mlnkmtlhpqqimigprfnralfdpllvvllalqllvvaglvraqtcpsvcscsnqfskv 60

Qy      61 ICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAF 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 icvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhleilqlsrnhirtieigaf 120

Qy     121 NGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 nglanlntlelfdnrlttipngafvylsklkelwlrnnpiesipsyafnripslrrldlg 180

Qy     181 ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQ 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 elkrlsyisegafeglsnlrylnlamcnlreipnltplikldeldlsgnhlsairpgsfq 240

Qy     241 GLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHH 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 glmhlqklwmiqsqiqviernafdnlqslveinlahnnltllphdlftplhhlerihlhh 300

Qy     301 NPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPP 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 npwncncdilwlswwikdmapsntaccarcntppnlkgryigeldqnyftcyapvivepp 360

Qy     361 ADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQD 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 adlnvtegmaaelkcrastsltsvswitpngtvmthgaykvriavlsdgtlnftnvtvqd 420

Qy     421 TGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTP 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 tgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmepsqdearttdnnvgptp 480

Qy     481 VVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGCFVAITLMA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 vvdwettnvttsltpqstrstektftipvtdinsgipgidevmkttkiiigcfvaitlma 540

Qy     541 AVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPMPAIEHEHLNHYNS 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 avmlvifykmrkqhhrqnhhaptrtveiinvddeitgdtpmeshlpmpaiehehlnhyns 600

Qy     601 YKSPFNHTTTVNTINSIHSSVHEPLLIRMNSKDNVQETQI 640
           ||||||||||||||||||||||||||||||||||||||||
Db     601 ykspfnhtttvntinsihssvhepllirmnskdnvqetqi 640


RESULT   8
AAB80262
ID   AAB80262 standard; Protein; 640 AA.
XX
AC   AAB80262;
XX
DT   24-APR-2001  (first entry)
XX
DE   Human PRO331 protein.
XX
KW   Human; PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;
KW   antiparkinsonian nootropic; neuroprotective; vulnerary; cardiant;
KW   antiangiogenic; vasotropic; antiasthmatic; antirheumatic; cancer;
KW   antiarthritic; antiinfertility; antidiabetic; antiviral; diabetes;
KW   ophthalmological; gene therapy; skin disease; gastrointestinal disorder;
KW   ischaemia; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200104311-A1.
XX
PD   18-JAN-2001.
XX
PF   22-FEB-2000; 2000WO-US04414.
XX
PR   07-JUL-1999;   99US-0143048.
PR   26-JUL-1999;   99US-0145698.
```

```
PR   28-JUL-1999;   99US-0146222.
PR   08-SEP-1999;   99WO-US20594.
PR   13-SEP-1999;   99WO-US20944.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   05-OCT-1999;   99WO-US23089.
PR   29-NOV-1999;   99WO-US28214.
PR   30-NOV-1999;   99WO-US28313.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   20-DEC-1999;   99WO-US30999.
PR   05-JAN-2000;   99WO-US00219.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Botstein D,  Desnoyers L,  Eaton DL,  Ferrara N;
PI   Filvaroff E,  Fong S,  Gao W,  Gerber H,  Gerritsen ME,  Goddard A;
PI   Godowski PJ,  Grimaldi CJ,  Gurney AL,  Hillan KJ,  Kljavin IJ;
PI   Mather JP,  Pan J,  Paoni NF,  Roy MA,  Stewart TA,  Tumas D;
PI   Williams PM,  Wood WI;
XX
DR   WPI; 2001-081051/09.
DR   N-PSDB; AAF72423.
XX
PT   Sixty one nucleic acids encoding PRO polypeptides which are useful in
PT   the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung
PT   squamous cell carcinoma) and neurodegenerative diseases (e.g.
PT   Alzheimer's disease) -
XX
PS   Claim 1; Fig 104; 393pp; English.
XX
CC   The present sequence is one of sixty one novel secreted and
CC   transmembrane PRO polypeptides. The PRO polypeptides are
CC   useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung
CC   squamous cell carcinoma), gastrointestinal disorders (e.g.
CC   enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,
CC   Parkinson's disease), wound repair, cardiovascular disorders (e.g.
CC   endometrial bleeding angiogenesis, ischaemias such as coronary
CC   ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,
CC   rheumatoid arthritis, multiple sclerosis), infertility, AIDS and
CC   diabetes and retinal disorders such as retinitis pigmentosum.
CC   The PRO nucleic acids have applications in molecular biology, including
CC   use as hybridization probes, and in chromosome and gene mapping.
XX
SQ   Sequence   640 AA;


  Query Match           100.0%;  Score 3362;  DB 22;  Length 640;
  Best Local Similarity  100.0%;  Pred. No. 2.3e-234;
  Matches  640;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MLNKMTLHPQQIMIGPRFNRALFDPLLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mlnkmtlhpqqimigprfnralfdpllvvllalqllvvaglvraqtcpsvcscsnqfskv 60

Qy      61 ICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAF 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 icvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhleilqlsrnhirtieigaf 120

Qy     121 NGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 nglanlntlelfdnrlttipngafvylsklkelwlrnnpiesipsyafnripslrrldlg 180

Qy     181 ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQ 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 elkrlsyisegafeglsnlrylnlamcnlreipnltplikldeldlsgnhlsairpgsfq 240

Qy     241 GLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHH 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 glmhlqklwmiqsqiqviernafdnlqslveinlahnnltllphdlftplhhlerihlhh 300

Qy     301 NPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPP 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 npwncncdilwlswwikdmapsntaccarcntppnlkgryigeldqnyftcyapvivepp 360

Qy     361 ADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQD 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 adlnvtegmaaelkcrastsltsvswitpngtvmthgaykvriavlsdgtlnftnvtvqd 420

Qy     421 TGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTP 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 tgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmepsqdearttdnnvgptp 480

Qy     481 VVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGCFVAITLMA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 vvdwettnvttsltpqstrstektftipvtdinsgipgidevmkttkiiigcfvaitlma 540

Qy     541 AVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPMPAIEHEHLNHYNS 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 avmlvifykmrkqhhrqnhhaptrtveiinvddeitgdtpmeshlpmpaiehehlnhyns 600

Qy     601 YKSPFNHTTTVNTINSIHSSVHEPLLIRMNSKDNVQETQI 640
           ||||||||||||||||||||||||||||||||||||||||
Db     601 ykspfnhtttvntinsihssvhepllirmnskdnvqetqi 640


RESULT   9
AAB65292
ID   AAB65292 standard; Protein; 640 AA.
XX
AC   AAB65292;
XX
DT   02-APR-2001  (first entry)
XX
DE   Human PRO331 protein sequence SEQ ID NO:501.
XX
KW   Human; secreted and transmembrane protein; PRO; cytostatic;
KW   cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
KW   diagnostic assay.
XX
OS   Homo sapiens.
XX
PN   WO200073454-A1.
XX
PD   07-DEC-2000.
XX
PF   30-MAR-2000; 2000WO-US08439.
XX
PR   02-JUN-1999;   99WO-US12252.
PR   23-JUN-1999;   99US-0141037.
PR   07-JUL-1999;   99US-0143048.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   28-JUL-1999;   99US-0146222.
PR   17-AUG-1999;   99US-0149396.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   08-OCT-1999;   99US-0158663.
PR   30-NOV-1999;   99WO-US28313.
PR   01-DEC-1999;   99WO-US28301.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   02-MAR-2000; 2000WO-US05841.
PR   15-MAR-2000; 2000WO-US06884.
PR   20-MAR-2000; 2000WO-US07377.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Botstein D,  Desnoyers L,  Eaton DL;
```

```
PI   Ferrara N,  Fong S,  Gerber H,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi CJ,  Gurney AL,  Kljavin IJ,  Napier MA,  Pan J,  Paoni NF;
PI   Roy MA,  Stewart TA,  Tumas D,  Watanabe CK,  Williams PM,  Wood WI;
PI   Zhang Z;
XX
DR   WPI; 2001-032160/04.
DR   N-PSDB; AAF44261.
XX
PT   PRO polynucleotides used to produce polypeptides used to target
PT   bioactive molecules such as toxins, radiolabels or antibodies, to
PT   specific cells, to cause targeted cell death -
XX
PS   Claim 12; Fig 314; 935pp; English.
XX
CC   The present invention describes human secreted and transmembrane PRO
CC   proteins. The PRO proteins have cytostatic activity. The PRO proteins
CC   can be used for targeted delivery of bioactive molecules, such as
CC   toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
CC   sequences, and their fragments, can be used as hybridisation probes, in
CC   chromosomal and gene mapping, and in the generation of anti-sense RNA
CC   and DNA. They may also be used to produce transgenic animals which are
CC   used to develop and screen therapeutically useful reagents. The PRO
CC   nucleotide and protein sequence can be used for tissue typing and in
CC   treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
CC   AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
CC   in the isolation of human PRO sequences. AAF44087 to AAF44269 and
CC   AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC   sequences given in the exemplification of the present invention.
XX
SQ   Sequence   640 AA;


  Query Match           100.0%;  Score 3362;  DB 22;  Length 640;
  Best Local Similarity  100.0%;  Pred. No. 2.3e-234;
  Matches  640;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MLNKMTLHPQQIMIGPRFNRALFDPLLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mlnkmtlhpqqimigprfnralfdpllvvllalqllvvaglvraqtcpsvcscsnqfskv 60

Qy      61 ICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAF 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 icvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhleilqlsrnhirtieigaf 120

Qy     121 NGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESISPSYAFNRIPSLRRLDLG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 nglanlntlelfdnrlttipngafvylsklkelwlrnnpiesipsyafnripslrrldlg 180

Qy     181 ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQ 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 elkrlsyisegafeglsnlrylnlamcnlreipnltplikldeldlsgnhlsairpgsfq 240

Qy     241 GLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHH 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 glmhlqklwmiqsqiqviernafdnlqslveinlahnnltllphdlftplhhlerihlhh 300

Qy     301 NPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPP 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 npwncncdilwlswwikdmapsntaccarcntppnlkgryigeldqnyftcyapvivepp 360

Qy     361 ADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQD 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 adlnvtegmaaelkcrastsltsvswitpngtvmthgaykvriavlsdgtlnftnvtvqd 420

Qy     421 TGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTP 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 tgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmepsqdearttdnnvgptp 480

Qy     481 VVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGCFVAITLMA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 vvdwettnvttsltpqstrstektftipvtdinsgipgidevmkttkiiigcfvaitlma 540

Qy     541 AVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPMPAIEHEHLNHYNS 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 avmlvifykmrkqhhrqnhhaptrtveiinvddeitgdtpmeshlpmpaiehehlnhyns 600

Qy     601 YKSPFNHTTTVNTINSIHSSVHEPLLIRMNSKDNVQETQI 640
           ||||||||||||||||||||||||||||||||||||||||
Db     601 ykspfnhtttvntinsihssvhepllirmnskdnvqetqi 640


RESULT  10
AAB53089
ID   AAB53089 standard; Protein; 640 AA.
XX
AC   AAB53089;
XX
DT   28-FEB-2001  (first entry)
XX
DE   Human angiogenesis-associated protein PRO331, SEQ ID NO:137.
XX
KW   Human; angiogenesis-associated protein; PRO; endothelial cell growth;
KW   cardiac hypertrophy; cardiovascular disorder; endothelial disorder;
KW   angiogenic disorder; atherosclerosis; osteoporosis; hypertension;
KW   myocardial infarction; diabetic retinopathy; rheumatoid arthritis;
KW   Crohn's disease; psoriasis; endometriosis; ulcer; wound healing; cancer;
KW   Alzheimer's disease; Huntington's disease; stroke; drug screening;
KW   gene therapy; transgenic animal.
XX
OS   Homo sapiens.
XX
PN   WO200053753-A2.
XX
PD   14-SEP-2000.
XX
PF   05-JAN-2000; 2000WO-US00219.
XX
PR   08-MAR-1999;   99WO-US05028.
PR   12-MAR-1999;   99US-0123957.
PR   14-MAY-1999;   99US-0134287.
PR   02-JUN-1999;   99WO-US12252.
PR   23-JUN-1999;   99US-0141037.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   01-SEP-1999;   99WO-US20111.
PR   08-SEP-1999;   99WO-US20594.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   05-OCT-1999;   99WO-US23089.
PR   30-NOV-1999;   99WO-US28313.
PR   30-NOV-1999;   99WO-US28409.
PR   02-DEC-1999;   99WO-US28564.
PR   02-DEC-1999;   99WO-US28565.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Ferrara N,  Gerber H,  Goddard A;
PI   Godowski PJ,  Gurney AL,  Hillan KJ,  Kuo SS,  Mark MR,  Marsters SA;
PI   Paoni NF,  Pitti RM,  Watanabe CK,  Williams PM,  Wood WI;
XX
DR   WPI; 2001-090793/10.
DR   N-PSDB; AAC97475.
XX
PT   New isolated nucleic acid for producing a PRO polypeptide, analyzing
PT   genetic disorders and treating cardiovascular, endothelial or
PT   angiogenic disorders, such as atherosclerosis, wounds or cancer -
XX
PS   Claim 69; Fig 52; 293pp; English.
XX
CC   The invention relates to novel human angiogenesis-associated proteins
CC   designated PRO proteins (AAB53064-B53097), and to nucleic acids encoding
CC   PRO proteins. The invention also relates to vectors and host cells
CC   comprising a PRO nucleic acid, the recombinant production of a PRO
CC   protein, PRO antibodies specific for a PRO protein, fusion proteins
CC   comprising a PRO protein, agonists or antagonists of a PRO protein, and
CC   compounds which inhibit the expression of a PRO gene. The invention
```

```
CC   additionally encompasses methods of identifying modulators of PRO
CC   expression or activity; diagnosing a cardiovascular, endothelial or
CC   angiogenic disorder, or a susceptibility to such a disorder by detecting
CC   mutations in a PRO gene, or the expression level of a PRO gene within a
CC   particular tissue; treating a cardiovascular, endothelial or angiogenic
CC   disorder via the administration of a PRO protein, PRO nucleic acid, or
CC   PRO agonist or antagonist; a retroviral gene therapy vector comprising a
CC   PRO nucleic acid; and methods of inhibiting or stimulating endothelial
CC   cell growth, cardiac hypertrophy or PRO-induced angiogenesis via the
CC   administration of a PRO protein, or an agonist or antagonist thereof.
CC   PRO nucleic acids, PRO proteins, antibodies against PRO proteins, PRO
CC   agonists and PRO antagonists may be used as therapeutic agents to treat
CC   cardiovascular, endothelial or angiogenic disorders, such as
CC   atherosclerosis, osteoporosis, myocardial infarction, hypertension,
CC   diabetic retinopathy, rheumatoid arthritis, Crohn's disease, psoriasis,
CC   endometriosis, ulcers, wounds, cancer, Alzheimer's disease, Huntington's
CC   disease, or stroke. PRO nucleic acids are additionally useful in the
CC   recombinant production of PRO proteins, as hybridisation probes to
CC   screen libraries to isolate cDNAs with sequence identity to PRO proteins,
CC   to map genes encoding PRO proteins, to analyse genetic disorders, and in
CC   gene therapy. PRO nucleic acids can also be used to produce transgenic
CC   animals useful for the development and screening of potential
CC   therapeutic agents. The present sequence represents a PRO protein of the
CC   invention.
XX
SQ   Sequence   640 AA;

  Query Match             100.0%;  Score 3362;  DB 22;  Length 640;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-234;
  Matches  640;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MLNKMTLHPQQIMIGPRFNRALFDPLLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mlnkmtlhpqqimigprfnralfdpllvvllalqllvvaglvraqtcpsvcscsnqfskv 60

Qy     61 ICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAF 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 icvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhleilqlsrnhirtieigaf 120

Qy    121 NGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 nglanlntlelfdnrlttipngafvylsklkelwlrnnpiesipsyafnripslrrldlg 180

Qy    181 ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQ 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 elkrlsyisegafeglsnlrylnlamcnlreipnltplikldeldlsgnhlsairpgsfq 240

Qy    241 GLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHH 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 glmhlqklwmiqsqiqviernafdnlqslveinlahnnltllphdlftplhhlerihlhh 300

Qy    301 NPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPP 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 npwncncdilwlswwikdmapsntaccarcntppnlkgryigeldqnyftcyapvivepp 360

Qy    361 ADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQD 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 adlnvtegmaaelkcrastsltsvswitpngtvmthgaykvriavlsdgtlnftnvtvqd 420

Qy    421 TGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTP 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 tgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmepsqdearttdnnvgptp 480

Qy    481 VVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGCFVAITLMA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 vvdwettnvttsltpqstrstektftipvtdinsgipgidevmkttkiiigcfvaitlma 540

Qy    541 AVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPMPAIEHEHLNHYNS 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 avmlvifykmrkqhhrqnhhaptrtveiinvddeitgdtpmeshlpmpaiehehlnhyns 600

Qy    601 YKSPFNHTTTVNTINSIHSSVHEPLLIRMNSKDNVQETQI 640
          |||||||||||||||||||||||||||||||||||||||
Db    601 ykspfnhtttvntinsihssvhepllirmnskdnvqetqi 640


RESULT  11
AAE13006
ID   AAE13006 standard; Protein; 713 AA.
XX
AC   AAE13006;
XX
DT   28-JAN-2002  (first entry)
XX
DE   Human leucine-rich repeat (LRR) family member protein.
XX
KW   Human; leucine-rich repeat; LRR; 31939 protein; therapy;
KW   cell proliferation; differentiation disorder; cancer; neuronal disorder;
KW   neurological disorder; demyelinating disease; multiple sclerosis;
KW   degenerative disease; Alzheimer's disease; Huntington's disease;
KW   spinocerebellar degeneration; nervous system; bone disorder;
KW   osteoporosis; immune disorder; rheumatoid arthritis; diabetes mellitus;
KW   cardiovascular disorder; liver disorder; viral disease; pain;
KW   metabolic disorder; chromosomal mapping; tissue typing; forensic biology;
KW   cytostatic; nootropic; neuroprotective; anticonvulsant; osteopathic;
KW   antirheumatic; antiarthritic; virucide; analgesic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..38
FT                   /label= Signal_peptide
FT   Protein         39..713
FT                   /note= "Mature human 31939 protein"
FT   Domain          56..85
FT                   /label= N-terminal_LRR_domain
FT   Domain          87..110
FT                   /label= LRR_domain
FT   Domain          111..134
FT                   /label= LRR_domain
FT   Domain          135..158
FT                   /label= LRR_domain
FT   Domain          159..182
FT                   /label= LRR_domain
FT   Domain          183..207
FT                   /label= LRR_domain
FT   Domain          208..229
FT                   /label= LRR_domain
FT   Domain          230..253
FT                   /label= LRR_domain
FT   Domain          254..277
FT                   /label= LRR_domain
FT   Domain          278..301
FT                   /label= LRR_domain
FT   Domain          311..362
FT                   /label= C-terminal_LRR_domain
FT   Domain          378..438
FT                   /label= Immunoglobulin_domain
XX
PN   WO200175105-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US10380.
XX
PR   31-MAR-2000; 2000US-193919P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Glucksmann MA;
XX
DR   WPI; 2001-662974/76.
DR   N-PSDB; AAD21287.
XX
PT   New polypeptide useful for treating cancer, multiple sclerosis,
```

```
PT   Alzheimer's disease, osteoporosis, arthritis, and metabolic and liver
PT   disorders comprises the 31939 polypeptide belonging to the leucine-rich
PT   repeat family  -
XX
PS   Claim 9; Fig 1A-1C; 117pp; English.
XX
CC   The present sequence is a human leucine-rich repeat (LRR) family
CC   member (referred as 31939) protein. The 31939 protein is useful for
CC   treating cell proliferative or differentiative disorder, e.g., cancer
CC   disorders or neuronal disorders, neurological disorders, demyelinating
CC   diseases such as multiple sclerosis, degenerative diseases such as
CC   Alzheimer's disease and Huntington's disease, spinocerebellar
CC   degenerations, disorders of the central or peripheral nervous system,
CC   bone disorders such as osteoporosis, immune disorders such as
CC   rheumatoid arthritis, diabetes mellitus, cardiovascular disorders,
CC   liver disorders, viral diseases, pain or metabolic disorders. They
CC   are also useful as markers of disorders or disease states, or for
CC   precursors or markers for the predisposition of disease states. They
CC   are useful in screening and detection assays such as chromosomal
CC   mapping, tissue typing, and forensic biology.
XX
SQ   Sequence   713 AA;


  Query Match             63.4%;  Score 2131.5;  DB 22;  Length 713;
  Best Local Similarity   61.6%;  Pred. No. 1.8e-145;
  Matches  409;  Conservative   81;  Mismatches  103;  Indels   71;  Gaps    9;

Qy      44 AQTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLE 103
           | :||  |||||| |:||| |::| |||  |  ||| ||| || ||:|: ::||||||||
Db      54 atscpvacscsnqasrvictrrdlaevpasipvntrylnlqengiqvirtdtfkhlrhle 113

Qy     104 ILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESI 163
           ||||||:| :| ||:|||||| :|||||||||||||:|  || |||||:|||||||||||
Db     114 ilqlsknlvrkievgafnglpslntlelfdnrlttvptqafeylsklrelwlrnnpiesi 173

Qy     164 PSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDE 223
           ||||||||:|||||||||||||| |||| ||||| ||||||| ||||::|||||  |::|:|
Db     174 psyafnrvpslrrldlgelkrleyiseaafeglvnlrylnlgmcnlkdipnltalvrlee 233

Qy     224 LDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLP 283
           |:|||| |  |||||||||  |:|||:: :|:  ||||||||:|:|| |:||:||||  ||
Db     234 lelsgnrldlirpgsfqgltslrklwlmhaqvatiernafddlksleelnlshnnlmslp 293

Qy     284 HDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGE 343
           ||||||| |||:||:||||:||||:||||||:|:  |||| |||||: |  ||||||||
Db     294 hdlftplhrlervhlnhnpwhcncdvlwlswwlketvpsnttccarchapaglkgryige 353

Qy     344 LDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRI 403
           |||::||||||||||| ||||||||||||||  ||:|||:|:|||||:||||:|:|||
Db     354 ldqshftcyapvivepptdlnvtegmaaelkcrtgtsmtsvnwltpngtlmthgsyrvri 413

Qy     404 AVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTP------------- 450
           :|| |||||||||||||| |||||:|| ||||||||||:|
Db     414 svlhdgtlnftnvtvqdtgqytcmvtnsagnttasatlnvsavdpvaaggtgsggggpgg 473

Qy     451 ----------FSYFSTVTVETME--PSQD--EARTTDNN-VGPTPVVDW----------E 485
                    ::||:||||||:|  | ::  : | |:    |||    |
Db     474 sggvgggsggytyfttvtvetletqpgeealqprgtekeppgpttdgvwgggrpgdaagp 533

Qy     486 TTNVTTSLTPQSTRSTEKTFTIPVTDI-NSGIPGIDEVMKTTKIIIGCFVAITLMAAVML 544
           :: ||:  |:|:| ||| ||:|:||:  : :  :|:||||||||||||||| ||||||
Db     534 asssttapaprssrptekaftvpitdvtenalkdlddvmkttkiiigcfvaitfmaavml 593

Qy     545 VIFYKMRKQHHRQNHHAPTRTVEIINVDDEIT---------------GDTPMESHLPMP 588
           | |||:||||    || ||||||||||:||:                |    :||| :|
Db     594 vafyklrkqhqlhkhhgptrtveiinvedelpaasavsvaaaaavasgggvggdshlalp 653

Qy     589 AIEHEHLNHYNSYKSPFNHTTTVNT------------INSIHSSVHEPLLIRMNSKDNVQ 636
           |:| :||||::     |    : |             :|||    ||||| :  ||:|||
Db     654 alerdhlnhhhyvadafkahyssnpsgggcggkgppglnsi----hepllfksgskenvq 709

Qy     637 ETQI 640
```

```
           ||||
Db     710 etqi 713


RESULT  12
AAB24073
ID   AAB24073 standard; Protein; 653 AA.
XX
AC   AAB24073;
XX
DT   29-JAN-2001  (first entry)
XX
DE   Human PRO1111 protein sequence SEQ ID NO:46.
XX
KW   Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;
KW   proliferation; tumourigenesis; identification; cancer; cytostatic;
KW   nootropic; neuroprotective; antiinflammatory; immunosuppressive;
KW   immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;
KW   neuronal disorder; glial disorder; astrocytal disorder; angiogenic;
KW   hypothalamic disorder; glandular disorder; macrophagal disorder;
KW   epithelial disorder; stromal disorder; blastocoelic disorder;
KW   inflammatory disorder; immunologic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200053755-A2.
XX
PD   14-SEP-2000.
XX
PF   06-JAN-2000; 2000WO-US00376.
XX
PR   08-MAR-1999;   99WO-US05028.
PR   02-JUN-1999;   99WO-US12252.
PR   23-JUN-1999;   99US-0141037.
PR   07-JUL-1999;   99US-0143048.
PR   26-JUL-1999;   99US-0145698.
PR   30-NOV-1999;   99WO-US28313.
PR   20-DEC-1999;   99WO-US30911.
PR   05-JAN-2000; 2000WO-US00219.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Goddard A,  Gurney AL,  Hillan KJ,  Roy MA;
PI   Watanabe CK,  Wood WI;
XX
DR   WPI; 2000-572270/53.
DR   N-PSDB; AAC58383.
XX
PT   Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
PT   treatment, diagnosis and prevention of cancer -
XX
PS   Claim 61; Fig 34; 286pp; English.
XX
CC   The present invention describes an isolated antibody  that binds to
CC   one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535,
CC   PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,
CC   PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,
CC   PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,
CC   PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
CC   growth. The PRO polypeptides and nucleotides are useful in the
CC   treatment, diagnosis and prevention of cancer. The antibodies and other
CC   anti-tumour compounds maybe used to treat various conditions, including
CC   those characterised by overexpression and/or activation of the amplified
CC   PRO genes. Exemplary conditions or disorders to be treated with such
CC   antibodies and other compounds include benign or malignant tumours
CC   (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
CC   colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
CC   carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
CC   leukaemias and lymphoid malignancies, other disorders such as neuronal,
CC   glial, astrocytal, hypothalamic and other glandular, macrophagal,
CC   epithelial, stromal and blastocoelic disorders, and inflammatory,
CC   angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
CC   primers and hybridisation probes used in the isolation of the human PRO
```

```
CC   sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human
CC   PRO polynucleotide and protein sequences given in the exemplification of
CC   the present invention.
XX
SQ   Sequence   653 AA;


  Query Match           55.5%;  Score 1865;  DB 21;  Length 653;
  Best Local Similarity  56.4%;  Pred. No. 2.8e-126;
  Matches  363;  Conservative  107;  Mismatches  130;  Indels   44;  Gaps   14;

Qy      29 VLLALQLLVVAGLVRA------QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLN 82
           | |  |: ::   : |      | ||||||||||||||:| |: | ||| || :||| ||
Db      22 vyltaqvwilcaaiaaaasagpqncpsvcscsnqfskvvctrrglsevpqgipsntryln 81

Qy      83 LHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNG 142
           | || ||:|: ::|:|| |||:||| || || ||:|||||||:||||||||| || ||:|
Db      82 lmenniqmiqadtfrhlhhlevlqlgrnsirqievgafnglaslntlelfdnwltvipsg 141

Qy     143 AFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYL 202
           || ||||:|||||||||||||||||:||| ||||||||:| |||||||||| ||:||
Db     142 afeylsklrelwlrnnpiesipsyafnrvpslmrldlgelkkleyisegafeglfnlkyl 201

Qy     203 NLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNA 262
           || |||:::::|||||: |:||::|||   |||||| ||  |:|||:: ||: :|||||
Db     202 nlgmcnikdmpnltplvgleelemsgnhfpeirpgsfhglsslkklwvmnsqvslierna 261

Qy     263 FDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPS 322
           || | ||||:|||||||: |||||||||| :|  :|||||||||:|||||:||:::  |:
Db     262 fdglaslvelnlahnnlsslphdlftplrylvelhlhhnpwncdcdilwlawwlreyipt 321

Qy     323 NTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLT 382
           |: || ||: | :::|||: |:||  | | || |:: | |||::||  |||||| :  :: \
Db     322 nstccgrchapmhmrgrylvevdqasfqcsapfimdaprdlnisegrmaelkcr-tppms 380

Qy     383 SVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLN 442
           || |: ||||::| :   ||:||:||||||::| : |||:|||||:|  ||: ||| ||
Db     381 svkwllpngtvlshasrhprisvlndgtlnfshvllsdtgvytcmvtnvagnsnasayln 440

Qy     443 VTAA--TTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTPVVDWETTNV------TTSLT 494
           |: |   |: :|:|:|:|||||| | | ::   |     | |     ||:       ||| |
Db     441 vstaelntsnysfftttvtvetteisped---ttrkykpvp-----ttstgyqpayttstt 492

Qy     495 P--QSTRSTEKTFTIPVTDINSGI-PGIDEVMKTTKIIIGCFVAITLMAAVMLVIFYKMR 551
             |:||   |   :| ||    :   :|||||||||||||||:||:|| ||::|||:|
Db     493 vliqttr-vpkqvavpatdttdkmqtsldevmkttkiiigcfvavtllaaamlivfyklr 551

Qy     552 KQHHRQNHHAPTRTVEIINVDDEITGDTPM----------ESHLPMPAIEHEHLNHYNS 600
           |:| :::     |||||| ||::|   |             |  : :| | |:|:| ||:
Db     552 krhqqrstvtaartveiiqvdedipaatsaaataapsgvsgegavvlpti-hdhin-ynt 609

Qy     601 YKSPFNHTTTVNTI-NSIH---SSVHEPLLIRMNSKDNVQETQI 640
           ||       | |:: ||:|   ::: || :|: ::|| ||||||
Db     610 ykpahgahwtenslgnslhptvttisepyiiqthtkdkvqetqi 653


RESULT  13
AAY66694
ID   AAY66694 standard; protein; 653 AA.
XX
AC   AAY66694;
XX
DT   05-APR-2000  (first entry)
XX
DE   Membrane-bound protein PRO1111.
XX
KW   Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW   pharmaceutical; receptor immunoadhesin; gene mapping.
XX
OS   Homo sapiens.
XX
PN   WO9963088-A2.
XX
PD   09-DEC-1999.
XX
PF   02-JUN-1999;   99WO-US12252.
XX
PR   02-JUN-1998;   98US-0087607.
PR   02-JUN-1998;   98US-0087609.
PR   02-JUN-1998;   98US-0087759.
PR   03-JUN-1998;   98US-0087827.
PR   04-JUN-1998;   98US-0088021.
PR   04-JUN-1998;   98US-0088025.
PR   04-JUN-1998;   98US-0088028.
PR   04-JUN-1998;   98US-0088029.
PR   04-JUN-1998;   98US-0088030.
PR   04-JUN-1998;   98US-0088033.
PR   04-JUN-1998;   98US-0088326.
PR   05-JUN-1998;   98US-0088167.
PR   05-JUN-1998;   98US-0088202.
PR   05-JUN-1998;   98US-0088212.
PR   05-JUN-1998;   98US-0088217.
PR   09-JUN-1998;   98US-0088655.
PR   10-JUN-1998;   98US-0088722.
PR   10-JUN-1998;   98US-0088730.
PR   10-JUN-1998;   98US-0088734.
PR   10-JUN-1998;   98US-0088738.
PR   10-JUN-1998;   98US-0088740.
PR   10-JUN-1998;   98US-0088741.
PR   10-JUN-1998;   98US-0088742.
PR   10-JUN-1998;   98US-0088810.
PR   10-JUN-1998;   98US-0088811.
PR   10-JUN-1998;   98US-0088824.
PR   10-JUN-1998;   98US-0088825.
PR   10-JUN-1998;   98US-0088826.
PR   11-JUN-1998;   98US-0088858.
PR   11-JUN-1998;   98US-0088861.
PR   11-JUN-1998;   98US-0088863.
PR   11-JUN-1998;   98US-0088876.
PR   12-JUN-1998;   98US-0089090.
PR   12-JUN-1998;   98US-0089105.
PR   16-JUN-1998;   98US-0089440.
PR   16-JUN-1998;   98US-0089512.
PR   16-JUN-1998;   98US-0089514.
PR   17-JUN-1998;   98US-0089532.
PR   17-JUN-1998;   98US-0089538.
PR   17-JUN-1998;   98US-0089598.
PR   17-JUN-1998;   98US-0089599.
PR   17-JUN-1998;   98US-0089600.
PR   17-JUN-1998;   98US-0089653.
PR   18-JUN-1998;   98US-0089801.
PR   18-JUN-1998;   98US-0089907.
PR   18-JUN-1998;   98US-0089908.
PR   19-JUN-1998;   98US-0089947.
PR   19-JUN-1998;   98US-0089948.
PR   19-JUN-1998;   98US-0089952.
PR   22-JUN-1998;   98US-0090246.
PR   22-JUN-1998;   98US-0090252.
PR   22-JUN-1998;   98US-0090254.
PR   23-JUN-1998;   98US-0090349.
PR   23-JUN-1998;   98US-0090355.
PR   24-JUN-1998;   98US-0090429.
PR   24-JUN-1998;   98US-0090431.
PR   24-JUN-1998;   98US-0090435.
PR   24-JUN-1998;   98US-0090444.
PR   24-JUN-1998;   98US-0090445.
PR   24-JUN-1998;   98US-0090461.
PR   24-JUN-1998;   98US-0090472.
PR   24-JUN-1998;   98US-0090535.
PR   24-JUN-1998;   98US-0090538.
PR   24-JUN-1998;   98US-0090540.
PR   24-JUN-1998;   98US-0090557.
PR   25-JUN-1998;   98US-0090676.
PR   25-JUN-1998;   98US-0090678.
PR   25-JUN-1998;   98US-0090688.
PR   25-JUN-1998;   98US-0090690.
```

PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 24-AUG-1998; 98US-0097661.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 26-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 99US-0115565.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX

DR WPI; 2000-072883/06.
DR N-PSDB; AAZ65033.
XX
PT Membrane-bound proteins and related nucleotide sequences -
XX
PS claim 12; Fig 157; 822pp; English.
XX
CC The invention provides membrane-bound PRO polypeptides and
CC polynucleotides encoding them. The PRO sequences of the invention were
CC identified based on extracellular domain homology screening. The PRO
CC sequences have homology with proteins including LDL receptors, TIE
CC ligands and various enzymes. The membrane-bound proteins and receptor
CC molecules are useful as pharmaceutical and diagnostic agents. Receptor
CC immunoadhesins, for instance, can be used as therapeutic agents to block
CC receptor-ligand interactions. The membrane-bound proteins can also be
CC employed for screening of potential peptide or small molecule inhibitors
CC of the relevant receptor/ligand interaction. The PRO encoding sequences
CC are useful as hybridization probes, in chromosome and gene mapping and in
CC the generation of antisense RNA and DNA. PRO nucleic acid sequences
CC will also be useful for the preparation of PRO polypeptides, especially
CC by recombinant techniques.
XX
SQ Sequence 653 AA;

Query Match 55.5%; Score 1865; DB 21; Length 653;
Best Local Similarity 56.4%; Pred. No. 2.8e-126;
Matches 363; Conservative 107; Mismatches 130; Indels 44; Gaps 14;

```
Qy     29 VLLALQLLVVAGLVRA------QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLN 82
           | |  |: ::   : |      | ||||||||||||||:| |: | ||| || :||| ||
Db     22 vyltaqvwilcaaiaaaasagpqncpsvcscsnqfskvvctrrglsevpqgipsntryln 81

Qy     83 LHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNG 142
          | || ||:|: ::|:|| |||:||| || || ||:|||||||:|||||||| || ||:|
Db     82 lmenniqmiqadtfrhlhhlevlqlgrnsirqievgafnglaslntlelfdnwltvipsg 141

Qy    143 AFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYL 202
          || |||||:|||||||||||||||||||:||| ||||||||:| |||||||||| ||:||
Db    142 afeylsklrelwlrnnpiesipsyafnrvpslmrldlgelkkleyisegafeglfnlkyl 201

Qy    203 NLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNA 262
          || |||:::|||||: |:||::|||   |||||| ||  |:|||:: ||: :|||||
Db    202 nlgmcnikdmpnltplvgleelemsgnhfpeirpgsfhglsslkklwvmnsqvslierna 261

Qy    263 FDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPS 322
          || | ||||:|||||||: |||||||||| :|  :|||||||||:||||||:||:::  |:
Db    262 fdglaslvelnlahnnlsslphdlftplrylvelhlhhnpwncdcdilwlawwlreyipt 321

Qy    323 NTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLT 382
          |: || ||: | :::|||: |:||  | | || |:: | |||::||  |||||| :  ::
Db    322 nstccgrchapmhmrgrylvevdqasfqcsapfimdaprdlnisegrmaelkcr-tppms 380

Qy    383 SVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLN 442
          || |: ||||||:: |   ||:||:||||||::| : |||:|||||:|  ||: ||| ||
Db    381 svkwllpngtvlshasrhprisvlndgtlnfshvllsdtgvytcmvtnvagnsnasayln 440

Qy    443 VTAA--TTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTPVVDWETTNV------TTSLT 494
          |: |   |: :|:|:|:|||||| | | ::   |     | |     ||:       ||| |
Db    441 vstaelntsnysffttvtvetteisped---ttrkykpvp-----ttstgyqpayttstt 492

Qy    495 P--QSTRSTEKTFTIPVTDINSGI-PGIDEVMKTTKIIIGCFVAITLMAAVMLVIFYKMR 551
            |:||   |   :| ||    :   :|||||||||||||||:||:|| ||::|||:|
Db    493 vliqttr-vpkqvavpatdttdkmqtsldevmkttkiiigcfvavtllaaamlivfyklr 551

Qy    552 KQHHRQNHHAPTRTVEIINVDDEITGDTPM----------ESHLPMPAIEHEHLNHYNS 600
          |:| :::     |||||| ||:::|   |             |  : :| | |:|:| ||:
Db    552 krhqqrstvtaartveiiqvdedipaatsaaataapsgvsgegavvlpti-hdhin-ynt 609

Qy    601 YKSPFNHTTTVNTI-NSIH---SSVHEPLLIRMNSKDNVQETQI 640
          ||       | |:: ||:|   ::: || :|: ::|| ||||||
Db    610 ykpahgahwtenslgnslhptvttisepyiiqthtkdkvqetqi 653
```

```
RESULT  14
AAE09438
ID   AAE09438 standard; Protein; 653 AA.
XX
AC   AAE09438;
XX
DT   19-NOV-2001  (first entry)
XX
DE   Human sbgPRO331a protein.
XX
KW   Human; Alzheimer's disease; amyotrophic lateral sclerosis;
KW   ALS; Zollinger-Ellison syndrome; immune system disease; schizophrenia;
KW   inflammation; haematopoietic disease; anxiety; feeding disorder; aging;
KW   anorexia; depression; cardiovascular disease; sleep disorder; seizure;
KW   memory alteration; migraine; stroke; asthma; neuropathy; hypoglycaemia;
KW   sexual disorder; growth abnormality; infection; autoimmune disease;
KW   rheumatoid arthritis; cataractogenesis; angiogenesis; atherosclerosis;
KW   cerebral ischaemia; cirrhosis; Huntington's disease; Hodgson's disease;
KW   hypercholesterolaemia; headache; amnesia; cardiac arrhythmia; obesity;
KW   diabetes mellitus; glomerulonephritis; renovascular hypertension;
KW   cancer; vaccine; gene therapy; sbgPRO331a gene.
XX
OS   Homo sapiens.
XX
PN   WO200160850-A1.
XX
PD   23-AUG-2001.
XX
PF   14-FEB-2001; 2001WO-US04703.
XX
PR   14-FEB-2000; 2000US-0182172.
PR   29-FEB-2000; 2000US-0186084.
PR   18-APR-2000; 2000US-0198583.
PR   04-OCT-2000; 2000US-0237963.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Agarwal P,  Kabnick KS,  Murdoch PR,  Rizvi SK,  Smith RF,  Xiang Z;
XX
DR   WPI; 2001-536566/59.
DR   N-PSDB; AAD16345.
XX
PT   New secreted and membrane associated polypeptides for treating
PT   Alzheimer's disease, psoriasis, cancer, enterocolitis, sleep and sexual
PT   disorders, stroke, and asthma  -
XX
PS   Claim 1; Page 56-58; 94pp; English.
XX
CC   The present sequence is a human sbgPRO331a protein,
CC   a membrane bound protein of the invention.
CC   The invention relates to secreted and membrane associated polypeptides
CC   and nucleic acid molecules encoding such polypeptides. Sequences of the
CC   invention are useful for treating diseases such as Alzheimer's disease,
CC   amyotrophic lateral sclerosis (ALS), Zollinger-Ellison syndrome, diseases
CC   of the immune system, haematopoietic disease, inflammation, anxiety,
CC   schizophrenia, feeding disorders, anorexia, depression, social, sexual
CC   and rewarded behaviour, cardiovascular disease, sleep disorder, learning
CC   and memory alteration and altered immune response, seizure, migraine,
CC   cancer, stroke, asthma, neuropathy, aging, sexual disorders, treatment
CC   of transsexuals, growth abnormalities, obesity, infections, autoimmune
CC   diseases (e.g. rheumatoid arthritis), cataractogenesis, angiogenesis,
CC   disorders associated with healthy maintenance of gastric mucosa and
CC   repair of acute and chronic mucosal lesion, lung carcinoma, cerebral
CC   ischaemia, atherosclerosis, cirrhosis, Huntington's disease, headache,
CC   amnesia, multiple sclerosis, Hodgson's disease, dilated cardiomyopathy,
CC   congestive heart failure, cardiac arrhythmias, hypercholesterolaemia,
CC   viral and non-viral hepatitis, type I and type II diabetes mellitus,
CC   glomerulonephritis, renovascular hypertension, hypoglycaemia, periodic
CC   paralyses, tendinitis and malignant hyperthermia. Polypeptides of the
CC   invention are used to identify membrane bound and soluble receptors.
CC   They are also useful as vaccines for inducing an immunological response
CC   in a mammal. Polynucleotides of the invention are used in gene therapy.
CC   They are also valuable for chromosome localisation studies and tissue
CC   expression studies.
XX
SQ   Sequence   653 AA;


  Query Match             55.5%;  Score 1865;  DB 22;  Length 653;
  Best Local Similarity   56.4%;  Pred. No. 2.8e-126;
  Matches  363;  Conservative  107;  Mismatches  130;  Indels   44;  Gaps   14;

Qy      29 VLLALQLLVVAGLVRA------QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLN 82
           | |  |: ::   : |      | ||||||||||||||:| |: | ||| || :||| ||
Db      22 vyltaqvwilcaaiaaaasagpqncpsvcscsnqfskvvctrrglsevpqgipsntryln 81

Qy      83 LHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNG 142
           | || ||:|: ::|:|| |||:||| || || ||:||||||:|||||||| || ||:|
Db      82 lmenniqmiqadtfrhlhhlevlqlgrnsirqievgafnglaslntlelfdnwltvipsg 141

Qy     143 AFVYLSKLKELWLRNNPIESISPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYL 202
           || |||||:|||||||||||||||||:||| |||||||:| |||||||||| ||:||
Db     142 afeylsklrelwlrnnpiesipsyafnrvpslmrldlgelkkleyisegafeglfnlkyl 201

Qy     203 NLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNA 262
           || |||:::::|||||: |:||::|||   |||||| ||  |:|||:: ||: :|||||
Db     202 nlgmcnikdmpnltplvgleelemsgnhfpeirpgsfhglsslkklwvmnsqvslierna 261

Qy     263 FDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPS 322
           || | ||||:|||||||: |||||||||  :|  :|||||||||:||||||:||:::  |:
Db     262 fdglaslvelnlahnnlsslphdlftplrylvelhlhhnpwncdcdilwlawwlreyipt 321

Qy     323 NTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLT 382
           |: || ||: | :::||||: |:||  | | || |:: | |||::||  ||||||  :  ::
Db     322 nstccgrchapmhmrgrylvevdqasfqcsapfimdaprdlnisegrmaelkcr-tppms 380

Qy     383 SVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLN 442
           || |: |||||::| :   ||:||:||||||::| : |||:|||||:|  ||: ||| ||
Db     381 svkwllpngtvlshasrhprisvlndgtlnfshvllsdtgvytcmvtnvagnsnasayln 440

Qy     443 VTAA--TTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTPVVDWETTNV------TTSLT 494
           |: |   |: :|:|:|||||| | | ::   |     | |     ||:       ||| |
Db     441 vstaelntsnysffttvtvetteisped---ttrkykpvp-----ttstgyqpayttstt 492

Qy     495 P--QSTRSTEKTFTIPVTDINSGI-PGIDEVMKTTKIIIGCFVAITLMAAVMLVIFYKMR 551
              |:||   |   :| ||    :   :|||||||||||||||:||:|| ||::|||:|
Db     493 vliqttr-vpkqvavpatdttdkmqtsldevmkttkiiigcfvavtllaaamlivfyklr 551

Qy     552 KQHHRQNHHAPTRTVEIINVDDEITGDTPM----------ESHLPMPAIEHEHLNHYNS 600
           |:| :::     |||||| ||::|   |             |  : :| | |:|:| ||:
Db     552 krhqqrstvtaartveiiqvdedipaatsaaataapsgvsgegavvlpti-hdhin-ynt 609

Qy     601 YKSPFNHTTTVNTI-NSIH---SSVHEPLLIRMNSKDNVQETQI 640
           ||       | |:: ||:|   ::: || :|: ::|| ||||||
Db     610 ykpahgahwtenslgnslhptvttisepyiiqthtkdkvqetqi 653


RESULT  15
AAU12390
ID   AAU12390 standard; Protein; 653 AA.
XX
AC   AAU12390;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human PRO1111 polypeptide sequence.
XX
KW   Human secretory and transmembrane; PRO; mammalian; cancer; lung;
KW   breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW   cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW   adipocyte; A-peptide; factor VIIA; gene therapy.
XX
OS   Homo sapiens.
XX
```

PN WO200140466-A2.
XX
PD 07-JUN-2001.
XX
PF 01-DEC-2000; 2000WO-US32678.
XX
PR 01-DEC-1999; 99WO-US28301.
PR 01-DEC-1999; 99WO-US28634.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28564.
PR 02-DEC-1999; 99WO-US28565.
PR 09-DEC-1999; 99US-0170262.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 20-DEC-1999; 99WO-US30999.
PR 30-DEC-1999; 99WO-US31243.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 24-FEB-2000; 2000WO-US05004.
PR 01-MAR-2000; 2000WO-US05601.
PR 20-MAR-2000; 2000WO-US07377.
PR 21-MAR-2000; 2000WO-US07532.
PR 30-MAR-2000; 2000WO-US08439.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 10-NOV-2000; 2000WO-US30873.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
DR WPI; 2001-408281/43.
DR N-PSDB; AAS21462.
XX
PT Isolated, secretory and transmembrane PRO polypeptide used to detect
PT other PRO polypeptides, link bioactive molecules to cells expressing
PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
PT lung, breast, prostate, cervical -
XX
PS Claim 12; Fig 438; 813pp; English.
XX
CC AAU12172-AAU12446 represent novel human secretory and transmembrane
CC PRO polypeptides. The PRO polypeptides are useful to detect other
CC PRO polypeptides, to link bioactive molecules to cells expressing
CC PRO polypeptides, to modulate biological activities of cells expressing
CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC polypeptide expression in a cell sample to that in a control sample.
CC Some of the 275 sequences are also useful to stimulate the release of
CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
CC proliferation or differentiation of chondrocytes, the proliferation or
CC gene expression in pericyte cells, the release of proteoglycans from
CC cartilage, the proliferation of inner ear utricular supporting cells or
CC of T-lymphocytes, the release of a cytokine from peripheral blood
CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
CC to factor VIIA. The PRO polypeptides can be used in assays to identify
CC molecules involved in binding interactions. The polynucleotides encoding
CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
CC transgenic or knock out animals and can be used in gene therapy.
XX
SQ Sequence 653 AA;

Query Match 55.5%; Score 1865; DB 22; Length 653;
Best Local Similarity 56.4%; Pred. No. 2.8e-126;
Matches 363; Conservative 107; Mismatches 130; Indels 44; Gaps 14;

```
Qy      29 VLLALQLLVVAGLVRA------QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLN 82
           | |  |: ::   : |      | ||||||||||||||:| |: | ||| || :||| ||
Db      22 vyltaqvwilcaaiaaaasagpqncpsvcscsnqfskvvctrrglsevpqgipsntryln 81

Qy      83 LHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNG 142
           | || ||:|: ::|:|| |||:||| || || ||:|||||||:|||||||| || ||:|
Db      82 lmenniqmiqadtfrhlhhlevlqlgrnsirqievgafnglaslntlelfdnwltvipsg 141

Qy     143 AFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYL 202
           || |||||:|||||||||||||||||:||| |||||||:| ||||||||| ||:||
Db     142 afeylsklrelwlrnnpiesipsyafnrvpslmrldlgelkkleyisegafeglfnlkyl 201

Qy     203 NLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNA 262
           || |||::::|||||: |:||::|||   |||||| ||  |:|||:: ||: :|||||
Db     202 nlgmcnikdmpnltplvgleelemsgnhfpeirpgsfhglsslkklwvmnsqvslierna 261

Qy     263 FDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPS 322
           || | ||||:|||||||: |||||||| :|  :||||||||:||||||:||::: |:
Db     262 fdglaslvelnlahnnlsslphdlftplrylvelhlhhnpwncdcdilwlawwlreyipt 321

Qy     323 NTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLT 382
           |: || ||: | :::|||: |:||  | | || |:: | |||::||  |||||| :  ::
Db     322 nstccgrchapmhmrgrylvevdqasfqcsapfimdaprdlnisegrmaelkcr-tppms 380

Qy     383 SVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLN 442
           || |: |||||::| :   ||:||:|||||::| : |||:|||||:|  ||: ||| ||
Db     381 svkwllpngtvlshasrhprisvlndgtlnfshvllsdtgvytcmvtnvagnsnasayln 440

Qy     443 VTAA--TTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTPVVDWETTNV------TTSLT 494
           |: |   |: :|:|:|||||| | | ::   |     | |     ||:       ||| |
Db     441 vstaelntsnysffttvtvetteisped---ttrkykpvp-----ttstgyqpayttstt 492

Qy     495 P--QSTRSTEKTFTIPVTDINSGI-PGIDEVMKTTKIIIGCFVAITLMAAVMLVIFYKMR 551
              |:||   |   :| ||    :   :|||||||||||||||:||:|| ||::|||:|
Db     493 vliqttr-vpkqvavpatdttdkmqtsldevmkttkiiigcfvavtllaaamlivfyklr 551

Qy     552 KQHHRQNHHAPTRTVEIINVDDEITGDTPM----------ESHLPMPAIEHEHLNHYNS 600
           |:| :::     |||||| ||::|   |             |  : :| | |:|:| ||:
Db     552 krhqqrstvtaartveiiqvdedipaatsaaataapsgvsgegavvlpti-hdhin-ynt 609

Qy     601 YKSPFNHTTTVNTI-NSIH---SSVHEPLLIRMNSKDNVQETQI 640
           ||       | |:: ||:|   ::: || :|: ::|| ||||||
Db     610 ykpahgahwtenslgnslhptvttisepyiiqthtkdkvqetqi 653
```

Search completed: August 26, 2002, 15:40:58
Job time: 206 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 2002, 15:39:27 ; Search time 20.98 Seconds
(without alignments)
2931.228 Million cell updates/sec

Title: US-09-905-056-292
Perfect score: 3362
Sequence: 1 MLNKMTLHPQQIMIGPRFNR.....VHEPLLIRMNSKDNVQETQI 640

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1376.5	40.9	421	2	T46266	hypothetical prote
2	417.5	12.4	1091	2	A58532	glial cell membran
3	410.5	12.2	361	2	A53860	chondroadherin pre
4	406	12.1	707	2	JC7763	neuronal leucine-r
5	399	11.9	1531	2	T42218	slit-1 protein hom
6	374.5	11.1	1523	2	T13953	MEGF5 protein - ra
7	350	10.4	1469	2	B36665	slit protein 2 pre
8	350	10.4	1480	2	A36665	slit protein 1 pre
9	348.5	10.4	789	2	T28714	hypothetical prote
10	348.5	10.4	1355	2	T28715	hypothetical prote
11	344.5	10.2	1025	2	T42626	secreted leucine-r
12	337.5	10.0	1535	2	S46224	peroxidasin - frui
13	335	10.0	605	2	JC5239	insulin-like growt
14	329.5	9.8	560	2	A60164	platelet membrane
15	320	9.5	605	2	A41915	insulin-like growt
16	312	9.3	603	2	JC1282	insulin-like growt
17	311	9.3	603	2	JC6128	insulin-like growt
18	302	9.0	359	1	NBHUC8	decorin precursor
19	296.5	8.8	594	2	T23841	hypothetical prote
20	296.5	8.8	626	1	NBHUIA	platelet glycoprot
21	293	8.7	357	2	S24317	decorin precursor
22	293	8.7	360	2	S06280	decorin precursor
23	293	8.7	360	2	I47020	decorin - rabbit
24	290	8.6	354	2	A55454	decorin precursor
25	289	8.6	907	2	JG0193	G protein-coupled
26	287.5	8.6	536	2	A34901	lysine carboxypept
27	287.5	8.6	610	2	T23836	hypothetical prote
28	287	8.5	354	2	S29145	decorin precursor
29	286	8.5	420	2	A53531	oncofetal trophobl
30	286	8.5	603	2	T24315	hypothetical prote
31	283.5	8.4	907	2	JE0176	orphan G protein-c
32	281.5	8.4	1389	2	T13852	gene wheeler prote
33	276.5	8.2	1385	2	T13887	tlr protein - frui
34	275.5	8.2	653	2	T25194	hypothetical prote
35	275.5	8.2	680	2	T19939	hypothetical prote
36	271.5	8.1	368	1	BGHUN	biglycan precursor
37	270.5	8.0	369	2	S20811	proteoglycan I - m
38	270.5	8.0	369	2	S32793	biglycan precursor
39	269.5	8.0	369	2	S32559	biglycan precursor
40	264.5	7.9	458	2	T19941	hypothetical prote
41	262	7.8	1066	2	T15864	hypothetical prote
42	258.5	7.7	562	2	T34319	hypothetical prote
43	257.5	7.7	1134	1	A29944	chaoptin precursor
44	252	7.5	575	2	T29972	hypothetical prote
45	250.5	7.5	382	2	I39068	proline- arginine-

ALIGNMENTS

RESULT 1
T46266
hypothetical protein DKFZp761A179.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T46266
R;Blum, H.; Bauersachs, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23034
A;Accession: T46266
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-421 <AAA>
A;Cross-references: EMBL:AL137451
A;Experimental source: adult amygdala; clone DKFZp761A179
C;Genetics:
A;Note: DKFZp761A179.1

Query Match 40.9%; Score 1376.5; DB 2; Length 421;
Best Local Similarity 62.6%; Pred. No. 8.5e-83;
Matches 256; Conservative 56; Mismatches 58; Indels 39; Gaps 6;

```
Qy    206 MCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDN 265
          ||||::|||||  |::|:||:|||| |  |||||||||  |:|||:: :|:  |||||||:
Db      1 MCNLKDIPNLTALVRLEELELSGNRLDLIRPGSFQGLTSLRKLWLMHAQVATIERNAFDD 60

Qy    266 LQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTA 325
          |:|| |:||:|||| ||||||||||| |||:||:||||:||||:||||||:|:  ||||
Db     61 LKSLEELNLSHNNLMSLPHDLFTPLHRLERVHLNHNPWHCNCDVLWLSWWLKETVPSNTT 120

Qy    326 CCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVS 385
          |||||: |  |||||||||||::|||||||||||| ||||||||||||||  ||:|||:
Db    121 CCARCHAPAGLKGRYIGELDQSHFTCYAPVIVEPPTDLNVTEGMAAELKCRTGTSMTSVN 180

Qy    386 WITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTA 445
          |:|||||:||||:|:|||:|| |||||||||||||| |||||:|| ||||||||||:|
Db    181 WLTPNGTLMTHGSYRVRISVLHDGTLNFTNVTVQDTGQYTCMVTNSAGNTTASATLNVSA 240

Qy    446 ATTTP----------------------FSYFSTVTVETME--PSQD--EARTTDNN-VG 477
                                       ::||:||||||:|  | ::  : | |:    |
Db    241 VDPVAAGGTGSGGGGPGGSGGVGGGSGGYTYFTTVTVETLETQPGEEALQPRGTEKEPPG 300

Qy    478 PTPVVDW----------ETTNVTTSLTPQSTRSTEKTFTIPVTDI-NSGIPGIDEVMKTT 526
          ||    |            :: ||:  |:|:| ||| ||:|:||:  : :  :|:|||||
Db    301 PTTDGVWGGGRPGDAAGPASSSTTAPAPRSSRPTEKAFTVPITDVTENALKDLDDVMKTT 360

Qy    527 KIIIGCFVAITLMAAVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDDEI 575
          |||||||||| |||||| |||:||||    || ||||||||||:||:
Db    361 KIIIGCFVAITFMAAVMLVAFYKLRKQHQLHKHHGPTRTVEIINVEDEL 409
```

```
RESULT   2
A58532
glial cell membrane glycoprotein LIG-1 precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 11-Apr-1997 #sequence_revision 11-Apr-1997 #text_change 05-Nov-1999
C;Accession: A58532
R;Suzuki, Y.; Sato, N.; Tohyama, M.; Wanaka, A.; Takagi, T.
J. Biol. Chem. 271, 22522-22527, 1996
A;Title: cDNA cloning of a novel membrane glycoprotein that is expressed specifically in
A;Reference number: A58532; MUID:96394313
A;Accession: A58532
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1091 <SUZ>
A;Cross-references: GB:D78572; NID:g1545806; PIDN:BAA11416.1; PID:g1545807
C;Superfamily: leucine-rich alpha-2-glycoprotein repeat homology; proteoglycan amino-ter
F;36-61/Domain: proteoglycan amino-terminal homology <PAH>
F;71-94/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>
F;95-117/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>
F;118-141/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
F;142-165/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>
F;166-189/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>
F;191-213/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>
F;214-237/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>
F;238-261/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>
F;262-285/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>
F;286-309/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR10>
F;310-333/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR11>
F;334-357/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR12>
F;358-381/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR13>
F;385-408/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR14>
F;409-432/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR15>
F;440-485/Domain: proteoglycan carboxyl-terminal homology <PCH>

  Query Match             12.4%;  Score 417.5;  DB 2;  Length 1091;
  Best Local Similarity   23.1%;  Pred. No. 2.1e-19;
  Matches  156;  Conservative   87;  Mismatches  208;  Indels  223;  Gaps   19;

Qy      81 LNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIP 140
           |:|: |:|::|:  :|: |  ||:|:| ||:|  :  ||| ||: :: | |  | |  :
Db     218 LDLNRNRIRLIEGLTFQGLDSLEVLRLQRNNISRLTDGAFWGLSKMHVLHLEYNSLVEVN 277

Qy     141 NGAFVYLSKLKELWLRNNPIESIPS------------YAFNRIPSLRRLDLGELKRL-- 185
           :|:   |: | :| | || |  |                 :|| :  |    | ||  |
Db     278 SGSLYGLTALHQLHLSNNSISRIQRDGWSFCQKLHELILSFNNLTRLDEESLAELSSLSI 337

Qy     186 --------SYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPG 237
                   |:|:||||:|| :|| |:                ||  ::||         |
Db     338 LRLSHNAISHIAEGAFKGLKSLRVLD----------------LDHNEISGTIEDT--SG 378

Qy     238 SFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIH 297
           :| || :| || :  ::|: : : ||  |:||  :||  | :  :  | |  : :|: ::
Db     379 AFTGLDNLSKLTLFGNKIKSVAKRAFSGLESLEHLNLGENAIRSVQFDAFAKMKNLKELY 438

Qy     298 LHHNPWNCNCDILWLSWWI--KDMAPSNTACCARCNTPPNLKGRYI-GELDQNY------ 348
           :    : |:| : ||  |:  : :    || ||    | :|||: |   |  ::
Db     439 ISSESFLCDCQLKWLPPWLMGRMLQAFVTATCAH---PESLKGQSIFSVLPDSFVCDDFP 495

Qy     349 --------------------FTCYA---------------------------------- 353
                              ||| |
Db     496 KPQIITQPETTMAVVGKDIRFTCSAASSSSSPMTFAWKKDNEVLANADMENFAHVRAQDG 555

Qy     354 ------------------------------------------PVIVEPPADLNVTEG 368
                                                     |   : | |: :  |
Db     556 EVMEYTTILHLRHVTFGHEGRYQCIITNHFGSTYSHKARLTVNVLPSFTKIPHDIAIRTG 615

Qy     369 MAAELKCRASTSLT-SVSWITPNGTVMTHGAYKVRIAVLSDGTLNF-TNVTVQDTGMYTC 426
             | |:| |:      ::|    ||     | : |: |: |  : | |:| : | |:|:|
Db     616 TTARLECAATGHPNPQIAWQKDGGTDFP-AARERRMHVMPDDDVFFITDVKIDDMGVYSC 674

Qy     427 MVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNV--------GP 478
              || |: :|:|:|| |              |     | :|   |    |         |

Db     675 TAQNSAGSVSANATLTVLE-----------TPSLAVPLEDRVVTVGETVAFQCKATGSP 722

Qy     479 TPVVDWETTNVTTSLT-----------------------------PQSTRSTEKTFTI 507
           || : |       |||                             |  |       :|
Db     723 TPRITWLKGGRPLSLTERHHFTPGNQLLVVQNVMIDDAGRYTCEMSNPLGTERAHSQLSI 782

Qy     508 PVTDINSGIPGIDEVMKTTKII-IGCFVAITLMAAVMLVIFYKMRKQHHRQNHHAPTRTV 566
             |      ||  :   |  |  |    :| | : | : | |: ||   :   :: | |
Db     783 LPT------PGCRKDGTTVGIFTIAVVCSIVLTSLVWVCIIYQTRK---KSEEYSVTNTD 833

Qy     567 EIINVDDEITGDTP 580
           | |     :  | |
Db     834 ETI-----VPPDVP 842


RESULT   3
A53860
chondroadherin precursor - bovine
N;Alternate names: 38K leucine-rich protein
C;Species: Bos primigenius taurus (cattle)
C;Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 05-Nov-1999
C;Accession: A53860
R;Neame, P.J.; Sommarin, Y.; Boynton, R.E.; Heinegard, D.
J. Biol. Chem. 269, 21547-21554, 1994
A;Title: The structure of a 38-kDa leucine-rich protein (chondroadherin) isolated fro
A;Reference number: A53860; MUID:94342341
A;Accession: A53860
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-361 <NEA>
A;Cross-references: GB:U08018; NID:g470671; PIDN:AAA21330.1; PID:g470672
C;Superfamily: leucine-rich alpha-2-glycoprotein repeat homology; proteoglycan carbox
C;Keywords: disulfide bond
F;300-346/Domain: proteoglycan carboxyl-terminal homology <PCH>

  Query Match             12.2%;  Score 410.5;  DB 2;  Length 361;
  Best Local Similarity   32.1%;  Pred. No. 1.4e-19;
  Matches  111;  Conservative   55;  Mismatches  145;  Indels   35;  Gaps    9;

Qy      28 VVLLALQLLVVAGLVRA-QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHEN 86
           ::||:| | ::| |: |    ||  | | :    ||| :  |:::|  :|  |:|||  |
Db      5 MLLLSLSLGLLASLLPALAACPQNCHCHSDLQHVICDKVGLQKIPK-VSEKTKLLNLQRN 63

Qy      87 QIQIIKVNSFK--------HLRHLEI----------------LQLSRNHIRTIEIGAFNG 122
           ::  |||:         ||:| :|                | || | || :  |||:
Db      64 NFPVLATNSFRAMPNLVSLHLQHCQIREVAAGAFRGLKQLIYLYLSHNDIRVLRAGAFDD 123

Qy     123 LANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGEL 182
           |  |  | |  |::| :| |    |  |  | | || |  : | ||     || | | |
Db     124 LTELTYLYLDHNKVTELPRGLLSPLVNLFILQLNNNKIRELRSGAFQGAKDLRWLYLSE- 182

Qy     183 KRLSYISEGAFEGLSNLRYLNLAMCNLREIPN--LTPLIKLDELDLSGNHLSAIRPGSFQ 240
           ||  :  || : : ||    |    |   |:  |: |  ::|| || | | :|   :||
Db     183 NSLSSLQPGALDDVENLAKFYLDRNQLSSYPSAALSKLRVVEELKLSHNPLKSIPDNAFQ 242

Qy     241 GL-MHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLH 299
             :|: ||:  : ::     ||  : :|  ::| :| |  || :   |   || : |
Db     243 SFGRYLETLWLDNTNLEKFSDGAFLGVTTLKHVHLENNRLHQLPSNF--PFDSLETLTLT 300

Qy     300 HNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELD 345
           :||| | | :  |  |::       | ||   :|   :|::| : |
Db     301 NNPWKCTCQLRGLRRWLEAKTSRPDATCA---SPAKFRGQHIRDTD 343


RESULT   4
JC7763
neuronal leucine-rich repeat protein-3 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C;Accession: JC7763
R;Fukamachi, K.; Matsuoka, Y.; Kitanaka, C.; Kuchino, Y.; Tsuda, H.
Biochem. Biophys. Res. Commun. 287, 257-263, 2001
```

```
A;Title: Rat neuronal leucine-rich repeat protein-3: Cloning and regulation of the gene
A;Reference number: JC7763; PMID:11549284
A;Contents: Fibrosarcoma cells
A;Accession: JC7763
A;Molecule type: mRNA
A;Residues: 1-707 <FUK>
A;Cross-references: GB:AF291437
C;Comment: This protein, a new member of the neuronal leucine-rich repeat protein family
in protein-protein interaction and functions as a cell adhesion molecule or soluble liga
C;Genetics:
A;Gene: nlrr-3
C;Keywords: cell adhesion

  Query Match            12.1%;  Score 406;  DB 2;  Length 707;
  Best Local Similarity  23.9%;  Pred. No. 6.8e-19;
  Matches  154;  Conservative   98;  Mismatches  239;  Indels  154;  Gaps   20;

Qy      31 LALQLLVVAGLVRAQTCPSVCSCS-----------NQFSKVICVRKNLREVPDGISTNTR 79
           ||:  || ||   :   || :|:|             : | | |    |   |  :  :|:
Db      14 LAITALVQAGDKKVD-CPQLCTCEIRPWFTPRSIYMEASTVDCNDLGLLNFPARLPADTQ 72

Qy      80 LLNLHENQIQIIK---------------VNSFKHLRHLEILQLSR--------NHIRTIE 116
           :| |  | |  |:                |:   : :: : ::|:        | :  :
Db      73 ILLLQTNNIARIEHSTDFPVNLTGLDLSQNNLSSVTNINVQKMSQLLSVYLEENKLTELP 132

Qy     117 IGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRR 176
             ||:||  | :  | |: |  |||| |  |  | | :| :: | |  |  :|:|
Db     133 EKCLYGLSNLQELYVNHNLLSAISPGAFVGLHNLLRLHLNSNRLQMINSKWFEALPNLEI 192

Qy     177 LDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIP--------------------NLT 216
           | ||:   :  | :  |: |  || | :|  || |:|                    |
Db     193 LMLGD-NPILRIKDMNFQPLLKLRSLVIAGINLTEVPDDALVGLENLESISFYDNRLNKV 251

Qy     217 PLIKLDE------LDLSGNHLSAIRPGSFQGLMHLQKLWM--------IQS--------- 253
           | : | :      |||: | :: || | |  ::||::| :        | |
Db     252 PQVALQKAVNLKFLDLNKNPINRIRRGDFSNMLHLKELGINNMPELVSIDSLAVDNLPDL 311

Qy     254 ---------QIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWN 304
                  ::  |  |||  |  |  : |  | |: | |     | :|: | : :| ||
Db     312 RKIEATNNPRLSYIHPNAFFRLPKLESLMLNSNALSALYHGTIESLPNLKEISIHSNPIR 371

Qy     305 CNCDILWLSW---WIKDMAPSNTACCARCNTPPNLKGRYIGELD-QNYFTCYAPVIVEP- 359
           |:| | |::     |: | | :  |      ||  :|: : ::  ::      |:|
Db     372 CDCVIRWINMNKTNIRFMEPDSLFCV----DPPEFQGQNVRQVHFRDMMEICLPLIAPES 427

Qy     360 -PADLNVTEGMAAELKCRASTS-LTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVT 417
            |: |:|       | |||:      : ||||:|  :     : :  | |:|||:   :|
Db     428 FPSILDVEADSYVSLHCRATAEPQPEIYWITPSGKRLLPNTLREKFYVHSEGTLDIRGIT 487

Qy     418 VQDTGMYTCMVSNSVGNTTAS-------------ATLNV----------------TAA 446
           :: |:|||: :| ||    |               :||:                :
Db     488 PKEGGLYTCIATNLVGADLKSIMIKVGGFVPQDNNGSLNIKIRDIRANSVLVSWKANSKI 547

Qy     447 TTTPFSYFSTVTVETMEPSQDEARTTD----NNVGPTPVVDW--------------ETT 487
             :   : : |  |  : :|     :|    |     |  ::                :
Db     548 LKSSVKWTAFVKTEDSQAAQSARIPSDVKVYNLTHLKPSTEYKICIDIPTIYQKSRKQCV 607

Qy     488 NVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGC 532
           ||||       :   |: |: |  :  |: ||  ||     : ||
Db     608 NVTTKSLEHDGKENGKSHTVFVACV-GGLLGIIGVM----CLFGC 647


RESULT   5
T42218
slit-1 protein homolog - rat
N;Alternate names: MEGF4 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C;Accession: T42218
R;Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A;Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A;Reference number: Z14126; MUID:98360089
A;Accession: T42218
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1531 <NAK>
A;Cross-references: EMBL:AB011530; NID:g3449289; PIDN:BAA32460.1; PID:g3449290
A;Experimental source: strain Sprague-Dawley; brain
C;Genetics:
A;Gene: MEGF4

  Query Match            11.9%;  Score 399;  DB 2;  Length 1531;
  Best Local Similarity  22.7%;  Pred. No. 5.4e-18;
  Matches  156;  Conservative  101;  Mismatches  235;  Indels  194;  Gaps   18;

Qy      31 LALQLLVVAGLVRAQTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQI 90
           | | |   |  : |  ||::|:|:   : | |    |: :|  |  ||  | |: | |
Db      18 LWLLLWAAAWRLGATACPALCTCTG--TTVDCHGTGLQAIPKNIPRNTERLELNGNNITR 75

Qy      91 IKVNSFKHLRHLEILQLSRNHIRTIEIGAFN---------------------------- 121
           |  | |  |: | :|||  | |  :| |||:
Db      76 IHKNDFAGLKQLRVLQLMENQIGAVERGAFDDMKELERLRLNRNQLQVLPELLFQNNQAL 135

Qy     122 -------------------GLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIES 162
                              |  :|  |:|  |::: |  |||  |  |: | | || | :
Db     136 SRLDLSENSLQAVPRKAFRGATDLKNLQLDKNQISCIEEGAFRALRGLEVLTLNNNNITT 195

Qy     163 IPSYAFNRIPSLR--RLDLGEL---KRLSYISE--------GAF---EGLSNLRYLNL-- 204
           ||  :|| :| ||  ||     |     |:::|:         | |    | ::|| ||:
Db     196 IPVSSFNHMPKLRTFRLHSNHLFCDCHLAWLSQWLRQRPTIGLFTQCSGPASLRGLNVAE 255

Qy     205 ----------------------------AMCN---------------------------- 208
                                       |||:
Db     256 VQKSEFSCSGQGEAAQVPACTLSSGSCPAMCSCSNGIVDCRGKGLTAIPANLPETMTEIR 315

Qy     209 -----LREIP--NLTPLIKLDELDLSGNHLSAIRPGSFQGLMH---------------- 244
                 :: ||    :|  ||  :||| | :: | | :||||
Db     316 LELNGIKSIPPGAFSPYRKLRRIDLSNNQIAEIAPDAFQGLRSLNSLVLYGNKITDLPRG 375

Qy     245 -------LQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIH 297
                   || | :  ::|  |  :|| :||:|  ::|  | :  |    || |  :: :|
Db     376 VFGGLYTLQLLLLNANKINCIRPDAFQDLQNLSLLSLYDNKIQSLAKGTFTSLRAIQTLH 435

Qy     298 LHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTC------ 351
           |  ||: |:|:: ||: :::       |   ||| :|  |  : ||::    | |
Db     436 LAQNPFICDCNLKWLADFLRTNPIETTG--ARCASPRRLANKRIGQIKSKKFRCSAKEQY 493

Qy     352 YAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWI---------TPNGT---------- 392
           : |   :   :  |  :|   |||   |:   | :         |  |
Db     494 FIPGTEDYHLNSECTSDVACPHKCRCEASVVECSGLKLSKIPERIPQSTTELRLNNNEIS 553

Qy     393 -VMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTPF 451
            :   | :| ::: |    |:   |:  : | :    | |  : ||:   :| :
Db     554 ILEATGLFK-KLSHLKKINLSNNKVSEIEDGTFEGATSVSELHLTANQLESVRSGMFRGL 612

Qy     452 SYFSTVTVETMEPS--QDEARTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPV 509
               |: :     |   ::: |   ||    : |    |  |:::| :  : :   |: :
Db     613 DGLRTLMLRNNRISCIHNDSFTGLRNVRLLSLYD----NHITTISPGAFDTLQALSTLNL 668

Qy     510 T----DINSGIPGIDEVMKTTKIIIG 531
               : |  :  : : ::  ||: |
Db     669 LANPFNCNCQLAWLGDWLRKRKIVTG 694


RESULT   6
T13953
MEGF5 protein - rat
N;Alternate names: slit protein homolog
C;Species: Rattus norvegicus (Norway rat)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C;Accession: T13953
R;Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
```

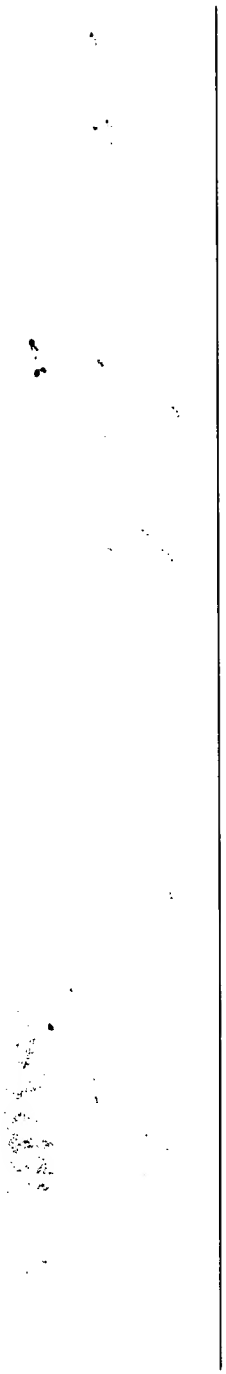
```
A;Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A;Reference number: Z14126; MUID:98360089
A;Accession: T13953
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1523 <NAK>
A;Cross-references: EMBL:AB011531; NID:g3449291; PIDN:BAA32461.1; PID:g3449292
C;Genetics:
A;Gene: MEGF5


  Query Match             11.1%;  Score 374.5;  DB 2;  Length 1523;
  Best Local Similarity   25.4%;  Pred. No. 2.2e-16;
  Matches  120;  Conservative   49;  Mismatches  147;  Indels  157;  Gaps    9;

Qy      31 LALQLLVVAGLVRAQTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQI 90
           |||  |  :      |  ||: |:||   : | |    || || ||  |   |:|  | |
Db      18 LALALASILSGPPAAACPTKCTCS--AASVDCHGLGLRAVPRGIPRNAERLDLDRNNITR 75

Qy      91 IKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKL 150
           |    |  |::| :| |  | :  || |||  |  |  | |  |:|  :|   |    ||
Db      76 ITKMDFTGLKNLRVLHLEDNQVSVIERGAFQDLKQLERLRLNKNKLQVLPELLFQSTPKL 135

Qy     151 KELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNL------ 204
             | |  | |: ||  ||  :  :: | | :   :| | :|||  | :|  | |
Db     136 TRLDLSENQIQGIPRKAFRGVTGVKNLQL-DNNHISCIEDGAFRALRDLEILTLNNNNIS 194

Qy     205 ------------------------------------------------------------ 204

Db     195 RILVTSFNHMPKIRTLRLHSNHLYCDCHLAWLSDWLRQRRTIGQFTLCMAPVHLRGFSVA 254

Qy     205 ----------------AMCN----------------------LREIP-NLTPLI----- 219
                            ||                        | ||| ||   |
Db     255 DVQKKEYVCPGPHSEAPACNANSLSCPSACSCSNNIVDCRGKGLTEIPANLPEGIVEIRL 314

Qy     220 ----------------KLDELDLSGNHLSAIRPGSFQ---------------------- 240
                           ||  :|:| | :| | | :||
Db     315 EQNSIKSIPAGAFIQYKKLKRIDISKNQISDIAPDAFQGLKSLTSLVLYGNKITEIPKGL 374

Qy     241 --GLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 298
             ||: || | :  ::|  :  | | :||:|  ::|  | |  :   || ||  :: :||
Db     375 FDGLVSLQLLLLNANKINCLRVNTFQDLQNLNLLSLYDNKLQTISKGLFAPLQSIQTLHL 434

Qy     299 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTC 351
             ||: |:| : ||: :::|  |  |:  |||::|  |  : | ::    | |
Db     435 AQNPFVCDCHLKWLADYLQD-NPIETS-GARCSSPRRLANKRISQIKSKKFRC 485


RESULT   7
B36665
slit protein 2 precursor - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change 19-May-2000
C;Accession: B36665
R;Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.; Artavanis-Tsakonas, S.
Genes Dev. 4, 2169-2187, 1990
A;Title: slit: an extracellular protein necessary for development of midline glia and co
A;Reference number: A36665; MUID:91099665
A;Accession: B36665
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1469 <ROT>
A;Cross-references: GB:X53959
C;Genetics:
A;Gene: FlyBase:sli
A;Cross-references: FlyBase:FBgn0003425
C;Superfamily: unassigned EGF-related proteins; EGF homology; leucine-rich alpha-2-glycc
F;66-91/Domain: proteoglycan amino-terminal homology <PAH1>
F;101-124/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>
F;125-148/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>
F;149-172/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
F;173-196/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>
F;197-220/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>
F;228-272/Domain: proteoglycan carboxyl-terminal homology <PCS1>
F;288-313/Domain: proteoglycan amino-terminal homology <PAH2>
F;323-346/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>
F;347-370/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>
F;371-394/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>
F;395-418/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>
F;419-442/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR10>
F;450-494/Domain: proteoglycan carboxyl-terminal homology <PCS2>
F;512-537/Domain: proteoglycan amino-terminal homology <PAH3>
F;547-571/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR11>
F;572-595/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR12>
F;596-619/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR13>
F;620-643/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR14>
F;651-695/Domain: proteoglycan carboxyl-terminal homology <PCS3>
F;708-733/Domain: proteoglycan amino-terminal homology <PAH4>
F;743-766/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR15>
F;767-790/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR16>
F;846-890/Domain: proteoglycan carboxyl-terminal homology <PCS4>
F;1028-1061/Domain: EGF homology <EGF>
F;1068-1099/Domain: EGF homology <EGF2>
F;1115-1148/Domain: EGF homology <EGF1>


  Query Match             10.4%;  Score 350;  DB 2;  Length 1469;
  Best Local Similarity   23.9%;  Pred. No. 8.4e-15;
  Matches  109;  Conservative   65;  Mismatches  152;  Indels  130;  Gaps    9;

Qy      46 TCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEIL 105
           :||  | |::    | |  |:|  ||  :  :|  : | :| |  :   ||   | |  :
Db     294 SCPHPCRCADGI--VDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKSFSSFRRLRRI 351

Qy     106 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 165
           || |:|  |   | :||  | || |: |::  :|:| |  |  |: | |  | |  |
Db     352 DLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLLNANEISCIRK 411

Qy     166 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM------CNLR--------- 210
           ||  : ||  | | :   :  :: | |: : ::: ::||       ||||
Db     412 DAFRDLHSLSLLSLYD-NNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRWLADYLHKN 470

Qy     211 ------------------------------------------------------------ 210

Db     471 PIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKLSGECRMDSDCPAMCHCEGTTV 530

Qy     211 --------EIPNLTPL--------------IKLD----------ELDLSGNHLSAIRPGS 238
                   |||   ||              |  |          :|:|  | |: | | :
Db     531 DCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKRNQLTGIEPNA 590

Qy     239 FQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 298
           |:|  |:|:| : :::|: |    |  |  |  :||  | :: :    |  |: |  ::|
Db     591 FEGASHIQELQLGENKIKEISNKMFLGLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNL 650

Qy     299 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVE 358
             ||:|||| : | :  ::  : :  |  |||  |  ::    | :|  : | | :
Db     651 ASNPFNCNCHLAWFAECVRKKSLNGGA--ARCGAPSKVRDVQIKDLPHSEFKCSS----- 703

Qy     359 PPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVM 394
                 :||   :  |  | : |        |||:
Db     704 -----ENSEGCLGDGYCPPSCTCT--------GTVV 726


RESULT   8
A36665
slit protein 1 precursor - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change 17-Nov-2000
C;Accession: A36665; A31640; S13523
R;Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.; Artavanis-Tsakonas, S.
Genes Dev. 4, 2169-2187, 1990
A;Title: slit: an extracellular protein necessary for development of midline glia and
A;Reference number: A36665; MUID:91099665
A;Accession: A36665
A;Status: preliminary
A;Molecule type: mRNA
```

```
A;Residues: 1-1480 <ROT>
A;Cross-references: GB:X53959; NID:g8614; PIDN:CAA37910.1; PID:g8615
R;Rothberg, J.M.; Hartley, D.A.; Walther, Z.; Artavanis-Tsakonas, S.
Cell 55, 1047-1059, 1988
A;Title: slit: An EGF-homologous locus of D. melanogaster involved in the development of
A;Reference number: A31640; MUID:89077533
A;Accession: A31640
A;Molecule type: DNA
A;Residues: 881-1182,'G',1185-1404,'GT',1463-1464,'YHA' <RO2>
A;Cross-references: GB:M23543; NID:g340939; PID:g514357
C;Genetics:
A;Gene: FlyBase:sli
A;Cross-references: FlyBase:FBgn0003425
A;Introns: 1351/3
C;Superfamily: unassigned EGF-related proteins; EGF homology; leucine-rich alpha-2-glycc
C;Keywords: alternative splicing; growth factor
F;66-91/Domain: proteoglycan amino-terminal homology <PAH1>
F;101-124/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>
F;125-148/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>
F;149-172/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
F;173-196/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>
F;197-220/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>
F;228-272/Domain: proteoglycan carboxyl-terminal homology <PCS1>
F;288-313/Domain: proteoglycan amino-terminal homology <PAH2>
F;323-346/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>
F;347-370/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>
F;371-394/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>
F;395-418/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>
F;419-442/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR10>
F;450-494/Domain: proteoglycan carboxyl-terminal homology <PCS2>
F;512-537/Domain: proteoglycan amino-terminal homology <PAH3>
F;547-571/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR11>
F;572-595/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR12>
F;596-619/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR13>
F;620-643/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR14>
F;651-695/Domain: proteoglycan carboxyl-terminal homology <PCS3>
F;708-733/Domain: proteoglycan amino-terminal homology <PAH4>
F;743-766/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR15>
F;767-790/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR16>
F;791-814/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR17>
F;815-838/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR18>
F;846-890/Domain: proteoglycan carboxyl-terminal homology <PCS4>
F;1028-1061/Domain: EGF homology <EGF>
F;1068-1099/Domain: EGF homology <EGF2>
F;1115-1148/Domain: EGF homology <EGF1>

  Query Match           10.4%;  Score 350;  DB 2;  Length 1480;
  Best Local Similarity  23.9%;  Pred. No. 8.5e-15;
  Matches  109;  Conservative   65;  Mismatches  152;  Indels  130;  Gaps    9;

Qy      46 TCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEIL 105
           :||  | |::    | |  |:|  ||  :  :|  : | :| |  :   ||   | |  :
Db     294 SCPHPCRCADGI--VDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKSFSSFRRLRRI 351

Qy     106 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 165
           || |:|  |    | :||  | || |: |::  :|:| |  |  |: | |  | |  |
Db     352 DLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLLNANEISCIRK 411

Qy     166 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM------CNLR--------- 210
           ||  : ||  | | :   :  :: | |: : ::: ::||       ||||
Db     412 DAFRDLHSLSLLSLYD-NNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRWLADYLHKN 470

Qy     211 ---------------------------------------------------------- 210

Db     471 PIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKLSGECRMDSDCPAMCHCEGTTV 530

Qy     211 --------EIPNLTPL--------------IKLD----------ELDLSGNHLSAIRPGS 238
                  |||   ||              |  |          :|:|  | |: | | :
Db     531 DCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKRNQLTGIEPNA 590

Qy     239 FQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 298
           |:|  |:|:| : :::|: |    |  |  |  :||  | :: :    |  |: |  ::|

Db     591 FEGASHIQELQLGENKIKEISNKMFLGLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNL 650

Qy     299 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVE 358
             ||:|||| : | :  ::  : :  |  |||  |  ::   | :|  : | | :
Db     651 ASNPFNCNCHLAWFAECVRKKSLNGGA--ARCGAPSKVRDVQIKDLPHSEFKCSS----- 703

Qy     359 PPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVM 394
                 :||   :  |  | : |        |||:
Db     704 -----ENSEGCLGDGYCPPSCTCT--------GTVV 726


RESULT   9
T28714
hypothetical protein T21D12.9a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T28714
R;Woessner, J.
submitted to the EMBL Data Library, August 1997
A;Description: The sequence of C. elegans cosmid T21D12.
A;Reference number: Z20514
A;Accession: T28714
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-789 <WOE>
A;Cross-references: EMBL:AF016687; PIDN:AAC48096.1; GSPDB:GN00022; CESP:T21D12.9a
A;Experimental source: strain Bristol N2; clone T21D12
C;Genetics:
A;Gene: CESP:T21D12.9a
A;Map position: 4
A;Introns: 38/2; 84/2; 132/2; 204/2; 275/2; 351/2; 519/3; 615/1; 714/1; 758/2

  Query Match           10.4%;  Score 348.5;  DB 2;  Length 789;
  Best Local Similarity  24.3%;  Pred. No. 4.6e-15;
  Matches  131;  Conservative   85;  Mismatches  233;  Indels   89;  Gaps   19;

Qy      75 STNTRL-LNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFD 133
           | || : | |  | |  :   ||  || || | |:|| || :   ||| | :|  : |
Db     101 SFNTLVTLKLARNHITTLNQFSFSRLRKLESLDLTRNMIREVRFLAFNQLPSLQNVSLAR 160

Qy     134 NRLTTIPNGAFVYLSKLKELWLRNNPIESIP------------SYAFNRI--------- 171
           | :  : :| |     || | |  | ::::                 ::|:|
Db     161 NDVYRLDDGMFYACEGLKHLNLSTNRVQAVTEGWMFGLTSLEVLDLSYNQIQSFHISSWS 220

Qy     172 --PSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPN--LTPLIKLDELDLS 227
             | |: | |    |:  :  |:|  |  |  | |:  ::  :     |  :  | :||||
Db     221 HTPKLKWLSL-HSNRIQSLPSGSFRVLRQLEELILSANSIDSLHKFALVGMSSLHKLDLS 279

Qy     228 GNHLS-AIRPGSF---QGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLP 283
           | |:  :  |:       :  |: |    :|::|| : ||:   :| |::|  | :  :
Db     280 SNTLAVCVEDGAVLYNTSMPFLRSLRFTNNQLRVIPKRAFERFPALEELDLTDNPIATIH 339

Qy     284 HDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGE 343
           : | ||  |:|: :: :   |:| | ||: ||  :     ::  |:|: || |   |:
Db     340 PEAFEPL-ELKRLVMNSSSILCDCQISWLASWIYRLKLDKSSIIAKCSYPPPLADLYVVA 398

Qy     344 LDQNYFTCY----APVIVEPPADLNVTEGMAAELKCRA-STSLTSVSW------------ 386
           :|    ||:        ||  | :::   |  |   |    |  |: |
Db     399 IDTANLTCHNDSPRAKIVRQPVEVSTLIGEKARFTCNVYGASPLSIEWRVMENGQPRVLV 458

Qy     387 ------ITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASAT 440
                 :: | | : :|  :  |   |:   |   || : |   | |:  |  | :  |
Db     459 QDSATFLSINRTAVVNGTFDER--ELAAAELLLDNVAMTDNSEYQCVARNRFG-SDFSTH 515

Qy     441 LNVTAATTTPFSYFSTVTVETM-----EPSQDEARTTDNNVGPTPVVDWETTNVTTSLTP 495
           : :       |:|   | | |     : ::     |    | | : |    |
Db     516 VKLQVYQAPKFTY----TPEDMPLLVGQTAKFLCAATGT---PRPEIKWAFEQI-----P 563

Qy     496 QSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGCF------VAITLMAAVMLVIF 547
                : :  |  |           :|  ||   | :      ||    |:  |::|
Db     564 FPAAEARRLYVTPNDD-------HIYIMNVTKEDQGAYTCHATNVAGQTQASANLIVF 614
```

```
RESULT  10
T28715
hypothetical protein T21D12.9b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T28715
R;Woessner, J.
submitted to the EMBL Data Library, August 1997
A;Description: The sequence of C. elegans cosmid T21D12.
A;Reference number: Z20514
A;Accession: T28715
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1355 <WOE>
A;Cross-references: EMBL:AF016687; PIDN:AAC48095.1; GSPDB:GN00022; CESP:T21D12.9b
A;Experimental source: strain Bristol N2; clone T21D12
C;Genetics:
A;Gene: CESP:T21D12.9b
A;Map position: 4
A;Introns: 38/2; 84/2; 132/2; 204/2; 275/2; 351/2; 519/3; 615/1; 714/1; 758/2; 786/3; 84

  Query Match             10.4%;  Score 348.5;  DB 2;  Length 1355;
  Best Local Similarity   24.3%;  Pred. No. 9.5e-15;
  Matches  131;  Conservative   85;  Mismatches  233;  Indels   89;  Gaps   19;

Qy      75 STNTRL-LNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFD 133
           | || : | |  | |  :   ||  || || | |:|| ||  :   ||| | :|  : |
Db     101 SFNTLVTLKLARNHITTLNQFSFSRLRKLESLDLTRNMIREVRFLAFNQLPSLQNVSLAR 160

Qy     134 NRLTTIPNGAFVYLSKLKELWLRNNPIESIP-------------SYAFNRI--------- 171
           |  :  : :| |     || | |  | ::::                 ::|:|
Db     161 NDVYRLDDGMFYACEGLKHLNLSTNRVQAVTEGWMFGLTSLEVLDLSYNQIQSFHISSWS 220

Qy     172 --PSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPN--LTPLIKLDELDLS 227
             | |: | |    |:  :  |:|  |  |  | |:  ::  :    |  :  | :||||
Db     221 HTPKLKWLSL-HSNRIQSLPSGSFRVLRQLEELILSANSIDSLHKFALVGMSSLHKLDLS 279

Qy     228 GNHLS-AIRPGSF---QGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLP 283
           | |:  :  |:       :  |: |    :|::|| : ||:   :| |::|  | :  :
Db     280 SNTLAVCVEDGAVLYNTSMPFLRSLRFTNNQLRVIPKRAFERFPALEELDLTDNPIATIH 339

Qy     284 HDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGE 343
           : | ||  |:|: :: :   |:| | ||: ||  :    ::  |:|: || |   |:
Db     340 PEAFEPL-ELKRLVMNSSSILCDCQISWLASWIYRLKLDKSSIIAKCSYPPPLADLYVVA 398

Qy     344 LDQNYFTCY----APVIVEPPADLNVTEGMAAELKCRA-STSLTSVSW------------ 386
           :|    ||:       ||  | :::   |  |    |     |  |: |
Db     399 IDTANLTCHNDSPRAKIVRQPVEVSTLIGEKARFTCNVYGASPLSIEWRVMENGQPRVLV 458

Qy     387 ------ITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASAT 440
                ::| | : :|  :  |   |:   |   || : |   | |:  |  | :  |
Db     459 QDSATFLSINRTAVVNGTFDER--ELAAAELLLDNVAMTDNSEYQCVARNRFG-SDFSTH 515

Qy     441 LNVTAATTTPFSYFSTVTVETM-----EPSQDEARTTDNNVGPTPVVDWETTNVTTSLTP 495
           : :       |:|    | | |     : ::     |     | | : |    :     |
Db     516 VKLQVYQAPKFTY----TPEDMPLLVGQTAKFLCAATGT---PRPEIKWAFEQI-----P 563

Qy     496 QSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGCF------VAITLMAAVMLVIF 547
           : :  | |          :|  ||   | :       ||    |:  |::|
Db     564 FPAAEARRLYVTPNDD-------HIYIMNVTKEDQGAYTCHATNVAGQTQASANLIVF 614


RESULT  11
T42626
secreted leucine-rich repeat-containing protein SLIT2 - mouse (fragment)
N;Alternate names: neurogenic extracellular slit protein
C;Species: Mus musculus (house mouse)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C;Accession: T42626
R;Holmes, G.P.; Negus, K.; Burridge, L.; Raman, S.; Algar, E.; Yamada, T.; Little, M.H.
Mech. Dev. 79, 57-72, 1998
A;Title: Distinct but overlapping expression patterns of two vertebrate slit homologs
A;Reference number: Z22177; MUID:99279238
A;Accession: T42626
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1025 <HOL>
A;Cross-references: EMBL:AF074960; NID:g4151258; PID:g4151259; PIDN:AAD04345.1
C;Genetics:
A;Gene: Slit2

  Query Match             10.2%;  Score 344.5;  DB 2;  Length 1025;
  Best Local Similarity   24.3%;  Pred. No. 1.2e-14;
  Matches  102;  Conservative   62;  Mismatches  149;  Indels  107;  Gaps    8;

Qy      47 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNS-FKHLRHLEIL 105
           ||  | |    : | |  : | ::|| |   |  | |: |:  :::   || |  | :
Db       2 CPEKCRCEG--TTVDCSNQRLNKIPDHIPQYTAELRLNNNEFTVLEATGIFKKLPQLRXI 59

Qy     106 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 165
             | | |  || ||| | : :| : |  |||  : :  |  |  || | ||:| |  : :
Db      60 NFSNNKITDIEEGAFEGASGVNEILLTSNRLENVQHKMFKGLESLKTLMLRSNRISCVGN 119

Qy     166 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM----CN------------ 208
           :|  : |:| | | :  ::: :: |||: | :|  |||      ||
Db     120 DSFIGLGSVRLLSLYD-NQITTVAPGAFDXLHSLSTLNLLANPFNCNCHLAWLGEWLRRK 178

Qy     209 --------------LREIP-----------------NLTPLIK---------------- 220
                         |:|||                 : :|| :
Db     179 RIVTGNPRCQKPYFLKEIPIQDVAIQDFTCDDGNDDNSCSPLSRCPSECTCLDTXVRCSN 238

Qy     221 ------------------------------------LDELDLSGNHLSAIRPGSFQGLMH 244
                                                |  :||| | :| :    |  :
Db     239 KGLKVLPKGIPKDVTELYLDGNQFTLVPKELSNYKHLTLIDLSNNRISTLSNQXFSNMTQ 298

Qy     245 LQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWN 304
           |  | :  :::: |    || |:||  ::|  |:::::|   |  |  |  : :  ||
Db     299 LLTLILSYNRLRCIPPRTFDGLKSLRLLSLHGNDISVVPEGAFNDLSALSHLAIGANPLY 358

Qy     305 CNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLN 364
           |:|:: ||| |:|  :        |||  |  :  : :        |||  |: :   |  |
Db     359 CDCNMQWLSDWVK--SEYKEPGIARCAGPGEMADKLLLTTPSKKFTCQGPMDITIQAKCN 416


RESULT  12
S46224
peroxidasin - fruit fly (Drosophila sp.)
C;Species: Drosophila sp.
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Oct-2000
C;Accession: S46224
R;Nelson, R.E.; Fessler, L.I.; Takagi, Y.; Blumberg, B.; Keene, D.R.; Olson, P.F.; Pa
EMBO J. 13, 3438-3447, 1994
A;Title: Peroxidasin: a novel enzyme-matrix protein of Drosophila development.
A;Reference number: S46224; MUID:94341255
A;Accession: S46224
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1535 <NEL>
A;Cross-references: GB:U11052; NID:g531384; PIDN:AAA61568.1; PID:g531385
C;Superfamily: peroxidasin; myeloperoxidase homology; proteoglycan amino-terminal hom
F;19-44/Domain: proteoglycan amino-terminal homology <PAH4>
F;661-1350/Domain: myeloperoxidase homology <MPX>

  Query Match             10.0%;  Score 337.5;  DB 2;  Length 1535;
  Best Local Similarity   25.6%;  Pred. No. 5.9e-14;
  Matches  106;  Conservative   58;  Mismatches  141;  Indels  109;  Gaps   13;

Qy      26 LLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHE 85
           |:: || | ||::|| |::  ||: |:|  :  | |:|  |  ||
Db       6 LMLQLLGL-LLLLAGGVQSVYCPAGCTCLER--TVRCIRAKLSAVP-------------- 48

Qy      86 NQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFV 145
```

```
                 |  :  : | |  |||  :   ||:||| | || | || |  : :||
Db     49 ----------KLPQDTQTLDLRFNHIEELPANAFSGLAQLTTLFLNDNELAYLQDGALN 97

Qy    146 YLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLA 205
          |: |: ::| || :  :|:  | |:| |               || |  | |
Db     98 GLTALRFVYLNNNRLSRLPATIFQRMPRL--------------EGIF--LEN------- 133

Qy    206 MCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDN 265
                                                :| : : :       |||
Db    134 ------------------------------------NDIWQLPAGL-------FDN 146

Qy    266 LQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHN--PWNCNCDI--LWLSWWIKDMAP 321
          |  |  : : :| || || | |  |::|:|: |  |    :||| :  ||  | : |:
Db    147 LPRLNRLIMYNNKLTQLPVDGFNRLNNLKRLRLDGNAIDIDCNCGVYSLWRRWHL-DVQR 205

Qy    322 SNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSL 381
             :   |  |  |: :    | :::| |  |  :  | |  |  |  |   || |   | |
Db    206 QLVSISLTCAAPQMLQNQGFSSLGEHHFKCAKPQFLVAPQDAQVAAGEQVELSCEV-TGL 264

Qy    382 --TSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVG 433
              ::|:  : |       : :  :|  |:|   :    | |:| |:  | :|
Db    265 HRPQITWM--HNTQELGLEEQTQAEILPSGSLLHRSADTSDMGIYQCIARNEMG 316


RESULT  13
JC5239
insulin-like growth factor acid-labile chain - baboon
C;Species: Papio sp. (baboon)
C;Date: 17-Apr-1997 #sequence_revision 09-May-1997 #text_change 09-May-1997
C;Accession: JC5239
R;Delhanty, P.; Baxter, R.C.
Biochem. Biophys. Res. Commun. 227, 897-902, 1996
A;Title: The cloning and expression of the baboon acid-labile subunit of the insulin-lik
A;Reference number: JC5239; MUID:97040714
A;Contents: liver
A;Accession: JC5239
A;Molecule type: mRNA
A;Residues: 1-605 <DEL>
C;Comment: This factor is structurally related to proinsulin and have insuline-like meta


  Query Match             10.0%;  Score 335;  DB 2;  Length 605;
  Best Local Similarity   20.8%;  Pred. No. 2.5e-14;
  Matches  125;  Conservative   64;  Mismatches  125;  Indels  286;  Gaps   13;

Qy     31 LALQLLVVA----------------GLVRAQTCPSVCSCS-----NQFSKVICVRKNLR 68
          ||| ||:::                |      ||: |:||     |: | | |  :||
Db      8 LALALLLLSWVALGPRSLEGAEPGTPGEAEGPACPATCACSYDDEVNELS-VFCSSRNLT 66

Qy     69 EVPDGISTNTRL---------------------------------------------- 80
           :||||   |:
Db     67 RLPDGIPGGTQALWLDSNNLSSIPPAAFRNLSSLAFLNLQGGQLGSLEPQALLGLENLCH 126

Qy     81 LNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIP 140
          |:|  ||:: : | :| :   | :| || | :  :| | | || ||  | |  | |  :|
Db    127 LHLERNQLRSLAVGTFAYTPALALLGLSNNRLSRLEDGLFEGLGNLWDLNLGWNSLAVLP 186

Qy    141 NGAF----------------VY--------LSKLKELWLRNNPIESIPSYAFNRIPSLRR 176
          : ||                |        |::|:|| |  | : :| :  | ::| |::
Db    187 DAAFRGLGGLRELVLAGNRLAYLQPALFSGLAELRELDLSRNALRAIKANVFAQLPRLQK 246

Qy    177 L------------------------------------------------------------ 177
          |
Db    247 LYLDRNLIAAVAPGAFLGLKALRWLDLSHNRVAGLLEDTFPGLLGLRVLRLSHNAIASLR 306

Qy    178 ----------------------------DLGELK-------RLSYISEGAFEGLSNLRY 201
                                       ||:|:       :|  :  ||| ||:|:
Db    307 PRTFEDLHFLEELQLGHNRIRQLAERSFEGLGQLEVLTLDHNQLQEVKVGAFLGLTNVAV 366

Qy    202 LNLAMCNLREIPN--LTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIE 259
          :||:   || :|      | ||  | | |: |  ||| :| ||  |::|::  : :  ||
Db    367 MNLSGNCLRNLPEQVFRGLGKLHSLHLEGSCLGRIRPHTFAGLSGLRRLFLKDNGLVGIE 426

Qy    260 RNAFDNLQSLVEINLAHNNLTLLPHDLFTPL---------------------------- 290
              :   | :|:|::|  | || ||| ||  |
Db    427 EQSLWGLAELLELDLTSNQLTHLPHQLFQGLGKLEYLLLSHNRLAELPADALGPLQRAFW 486

Qy    291 ----HH-----------------------------------LERIHLHHNPWNCNCDILW 311
              |:                                   |||: |  |||:|:| :
Db    487 LDVSHNRLEALPGSLLASLGRLRYLNLRNNSLRTFTPQPPGLERLWLEGNPWDCSCPLKA 546

Qy    312 LSWWIKDMAPSNTACCAR----------------------CNTPPNLKGRYIGELDQNYF 349
          |    :| |  | :  |                      | :|| : |  : :| : :|
Db    547 L----RDFALQNPSAVPRFVQAICEGDDCQPPVYTYNNITCASPPEVAGLDLRDLGEAHF 602


RESULT  14
A60164
platelet membrane glycoprotein V precursor - human
C;Species: Homo sapiens (man)
C;Date: 12-Jan-1993 #sequence_revision 24-Feb-1994 #text_change 05-Nov-1999
C;Accession: A48030; A60164; A35483; B35483; C35483; A60432; A47507; S34329
R;Lanza, F.; Morales, M.; de La Salle, C.; Cazenave, J.P.; Clemetson, K.J.; Shimomura
J. Biol. Chem. 268, 20801-20807, 1993
A;Title: Cloning and characterization of the gene encoding the human platelet glycopr
A;Reference number: A48030; MUID:94012616
A;Accession: A48030
A;Molecule type: DNA
A;Residues: 1-560 <LA2>
A;Cross-references: EMBL:Z23091; NID:g312501; PIDN:CAA80637.1; PID:g312502
R;Shimomura, T.; Fujimura, K.; Maehama, S.; Takemoto, M.; Oda, K.; Fujimoto, T.; Oyam
Blood 75, 2349-2356, 1990
A;Title: Rapid purification and characterization of human platelet glycoprotein V: th
A;Reference number: A60164; MUID:90275263
A;Accession: A60164
A;Molecule type: protein
A;Residues: 365-384,'X',386-390,'X',392-395,'X',397;188-208,'I',210;27-50,'X',52-53;1
,'XX',108,'T';61-72,'TK',75-77;'V',56-57;'G',479-487,'X',489-498,'X',500,'X',502-503,
R;Roth, G.J.; Church, T.A.; McMullen, B.A.; Williams, S.A.
Biochem. Biophys. Res. Commun. 170, 153-161, 1990
A;Title: Human platelet glycoprotein V: a surface leucine-rich glycoprotein related t
A;Reference number: A35483; MUID:90321220
A;Accession: A35483
A;Molecule type: protein
A;Residues: 145-166,'I',168-169,'X',171-172 <ROT>
A;Note: this proteolytic fragment was designated peptide M392
A;Accession: B35483
A;Molecule type: protein
A;Residues: 121-129,'W',131-135;466-468,'X',470 <RO2>
A;Note: this material was designated peptide M393 but may contain two peptides
A;Accession: C35483
A;Molecule type: protein
A;Residues: 252-266,'H',268-272,'X',274-279,'I',281-284,'I',286 <RO3>
A;Note: this proteolytic fragment was designated peptide M401
R;Zafar, R.S.; Walz, D.A.
Thromb. Res. 53, 31-44, 1989
A;Title: Platelet membrane glycoprotein V: characterization of the thrombin-sensitive
A;Reference number: A60432; MUID:89162331
A;Accession: A60432
A;Molecule type: protein
A;Residues: 477-478,'FX',481-485,'E',487,'V',489-492,'NQ',495,'E',497-498 <ZAF>
R;Hickey, M.J.; Hagen, F.S.; Yagi, M.; Roth, G.J.
Proc. Natl. Acad. Sci. U.S.A. 90, 8327-8331, 1993
A;Title: Human platelet glycoprotein V: characterization of the polypeptide and the r
A;Reference number: A47507; MUID:93391348
A;Accession: A47507
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-560 <RES>
A;Cross-references: GB:L11238; NID:g388759; PIDN:AAA03069.1; PID:g388760
C;Comment: This platelet membrane protein is a substrate for thrombin.
C;Comment: The amino end of the intact protein is blocked.
C;Comment: This protein is absent in Bernard-Soulier syndrome.
C;Genetics:
A;Gene: GDB:GP5
A;Cross-references: GDB:230236; OMIM:173511
```

```
A;Map position: 5pter-5qter
C;Superfamily: leucine-rich alpha-2-glycoprotein repeat homology
C;Keywords: blocked amino end; glycoprotein; platelet; tandem repeat; transmembrane prot

  Query Match             9.8%;  Score 329.5;  DB 2;  Length 560;
  Best Local Similarity   29.7%;  Pred. No. 5.2e-14;
  Matches   83;  Conservative   43;  Mismatches  104;  Indels   49;  Gaps    3;

Qy      66 NLREVPDG-----------------ISTNTRLLN---------LHENQIQIIKVNSFKH 98
           ||  :| |                   :| :: |||         | | |: |   :| |
Db     181 NLTHLPKGLLGAQAKLERLLLHSNRLVSLDSGLLNSLGALTELQFHRNHIRSIAPGAFDR 240

Qy      99 LRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNN 158
           | :|  | |||||:  :    |    ||  | ||:| |  :|   |  :  |:|||
Db     241 LPNLSSLTLSRNHLAFLPSALFLHSHNLTLLTLFENPLAELPGVLFGEMGGLQELWLNRT 300

Qy     159 PIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPL 218
           : ::|: ||  :  || | :    ||| : :|||:||  |:
Db     301 QLRTLPAAAFRNLSRLRYLGVTLSPRLSALPQGAFQGLGELQV----------------- 343

Qy     219 IKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNN 278
           | |  | |:|:  |  :||  |::: : ::::: : |  | || ||  : | ||
Db     344 -----LALHSNGLTALPDGLLRGLGKLRQVSLRRNRLRALPRALFRNLSSLESVQLDHNQ 398

Qy     279 LTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIK 317
           |  || |:|  |  |  : | || | |:| :     |::
Db     399 LETLPGDVFGALPRLTEVLLGHNSWRCDCGLGPFLGWLR 437


RESULT  15
A41915
insulin-like growth factor-binding complex acid-labile chain precursor - human
N;Alternate names: Acid-Labile Subunit (ALS)
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Nov-1999
C;Accession: A41915
R;Leong, S.R.; Baxter, R.C.; Camerato, T.; Dai, J.; Wood, W.I.
Mol. Endocrinol. 6, 870-876, 1992
A;Title: Structure and functional expression of the acid-labile subunit of the insulin-l
A;Reference number: A41915; MUID:92357025
A;Accession: A41915
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-605 <LEO>
A;Cross-references: GB:M86826; NID:g184807; PIDN:AAA36047.1; PID:g184808
A;Experimental source: liver
A;Note: sequence extracted from NCBI backbone (NCBIP:110171)
C;Superfamily: leucine-rich alpha-2-glycoprotein repeat homology
F;75-98/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>
F;99-122/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>
F;123-146/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
F;147-170/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>
F;171-194/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>
F;195-218/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>
F;219-242/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>
F;243-266/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>
F;267-290/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>
F;291-314/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR10>
F;315-338/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR11>
F;339-362/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR12>
F;363-386/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR13>
F;387-410/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR14>
F;411-434/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR15>
F;435-458/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR16>
F;459-482/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR17>
F;483-506/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR18>
F;507-529/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR19>

  Query Match             9.5%;  Score 320;  DB 2;  Length 605;
  Best Local Similarity   29.8%;  Pred. No. 2.4e-13;
  Matches   95;  Conservative   50;  Mismatches  124;  Indels   50;  Gaps    6;

Qy      31 LALQLLVVA----------------GLVRAQTCPSVCSCS-----NQFSKVICVRKNLR 68
           ||| ||:::                |      ||: | ||     :: | | |  :||
Db       8 LALALLLLSWVALGPRSLEGADPGTPGEAEGPACPAACVCSYDDDADELS-VFCSSRNLT 66

Qy      69 EVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI---------------------- 104
           :|||:   |: | | |  | :  :   :|::|  |
Db      67 RLPDGVPGGTQALWLDGNNLSSVPPAAFQNLSSLGFLNLQGGQLGSLEPQALLGLENLCH 126

Qy     105 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 164
           | | || :|:: :| |     | :| | :|||: : :| |  |  | :| |  | :  :|
Db     127 LHLERNQLRSLALGTFAHTPALASLGLSNNRLSRLEDGLFEGLGSLWDLNLGWNSLAVLP 186

Qy     165 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPN--LTPLIKLD 222
           ||  : ||| | |    ||:|:    | ||: || |:|:   || |       | :|
Db     187 DAAFRGLGSLRELVLAG-NRLAYLQPALFSGLAELRELDLSRNALRAIKANVFVQLPRLQ 245

Qy     223 ELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLL 282
           :| |  | ::|: ||:| ||  |: | :  :::  :  : |  |  |  : |:|| :  |
Db     246 KLYLDRNLIAAVAPGAFLGLKALRWLDLSHNRVAGLLEDTFPGLLGLRVLRLSHNAIASL 305

Qy     283 PHDLFTPLHHLERIHLHHN 301
           |  || || : | ||
Db     306 RPRTFKDLHFLEELQLGHN 324


Search completed: August 26, 2002, 15:41:49
Job time: 142 sec
```



GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 2002, 15:41:28 ; Search time 13.57 Seconds
(without alignments)
1826.125 Million cell updates/sec

Title: US-09-905-056-292
Perfect score: 3362
Sequence: 1 MLNKMTLHPQQIMIGPRFNR.....VHEPLLIRMNSKDNVQETQI 640

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	438.5	13.0	713	1	GAC1_HUMAN	O75325 homo sapien
2	411.5	12.2	361	1	CHAD_BOVIN	Q27972 bos taurus
3	369.5	11.0	567	1	GPV_MOUSE	O08742 mus musculu
4	361.5	10.8	567	1	GPV_RAT	O08770 rattus norv
5	350	10.4	1480	1	SLIT_DROME	P24014 drosophila
6	335	10.0	605	1	ALS_PAPHA	O02833 papio hamad
7	329.5	9.8	560	1	GPV_HUMAN	P40197 homo sapien
8	320	9.5	605	1	ALS_HUMAN	P35858 homo sapien
9	312	9.3	603	1	ALS_RAT	P35859 rattus norv
10	311	9.3	603	1	ALS_MOUSE	P70389 mus musculu
11	302	9.0	359	1	PGS2_HUMAN	P07585 homo sapien
12	299	8.9	360	1	PGS2_HORSE	O46542 equus cabal
13	296.5	8.8	626	1	GPBA_HUMAN	P07359 homo sapien
14	295	8.8	360	1	PGS2_CANFA	Q29393 canis famil
15	295	8.8	360	1	PGS2_PIG	Q9xsd9 sus scrofa
16	293	8.7	357	1	PGS2_CHICK	P28675 gallus gall
17	293	8.7	360	1	PGS2_BOVIN	P21793 bos taurus
18	293	8.7	360	1	PGS2_RABIT	Q28888 oryctolagus
19	290	8.6	354	1	PGS2_MOUSE	P28654 mus musculu
20	287.5	8.6	536	1	CBP8_HUMAN	P22792 homo sapien
21	287	8.5	354	1	PGS2_RAT	Q01129 rattus norv
22	276	8.2	353	1	KERA_COTJA	Q9de66 coturnix co
23	276	8.2	966	1	Y918_HUMAN	O94991 homo sapien
24	274.5	8.2	331	1	PLIB_AGKBL	O93233 agkistrodon
25	274	8.1	353	1	KERA_CHICK	O42235 gallus gall
26	271.5	8.1	368	1	PGS1_HUMAN	P21810 homo sapien
27	271.5	8.1	369	1	PGS1_CANFA	O02678 canis famil
28	271.5	8.1	372	1	PGS1_HORSE	O46403 equus cabal
29	270.5	8.0	369	1	PGS1_MOUSE	P28653 mus musculu
30	270.5	8.0	369	1	PGS1_RAT	P47853 rattus norv
31	269.5	8.0	369	1	PGS1_SHEEP	O46390 ovis aries
32	268.5	8.0	369	1	PGS1_BOVIN	P21809 bos taurus
33	268	8.0	423	1	OMD_RAT	Q9z1s7 rattus norv
34	264	7.9	782	1	CHAO_TRICA	P82963 tribolium c
35	258.5	7.7	905	1	TLR3_MOUSE	Q99mb1 mus musculu
36	256	7.6	1315	1	CHAO_DROME	P12024 drosophila
37	254.5	7.6	1049	1	TLR7_HUMAN	Q9nyk1 homo sapien
38	253	7.5	977	1	Y848_HUMAN	O94933 homo sapien
39	250.5	7.5	382	1	PRLP_HUMAN	P51888 homo sapien
40	249	7.4	904	1	TLR3_HUMAN	O15455 homo sapien
41	249	7.4	1050	1	TLR7_MOUSE	P58681 mus musculu
42	248	7.4	352	1	KERA_BOVIN	O62702 bos taurus
43	247	7.3	423	1	OMD_MOUSE	O35103 mus musculu
44	246	7.3	352	1	KERA_HUMAN	O60938 homo sapien
45	245.5	7.3	440	1	OMGP_HUMAN	P23515 homo sapien

ALIGNMENTS

RESULT 1
GAC1_HUMAN
ID GAC1_HUMAN STANDARD; PRT; 713 AA.
AC O75325;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Glioma amplified on chromosome 1 protein precursor.
GN GAC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Glial tumor;
RX MEDLINE=98324709; PubMed=9662332;
RA Malfoy B., Almeida A., Zhu X.X., Vogt N., Tyagi R., Muleris M.,
RA Dutrillaux A.-M., Dutrillaux B., Ross D., Hanash S.;
RT "GAC1, a new member of the leucine-rich repeat superfamily on
RT chromosome band 1q32.1, is amplified and overexpressed in malignant
RT gliomas.";
RL Oncogene 16:2997-3002(1998).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -!- TISSUE SPECIFICITY: OVERAMPLIFIED IN MALIGNANT GLIOMAS.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 11 LEUCINE-RICH REPEATS (LRR).
CC ----------

CC ----------
DR EMBL; AF030435; AAC39792.1; -.
DR MIM; 605492; -.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR003591; LRR_typ.
DR Pfam; PF00047; ig; 1.
DR Pfam; PF00560; LRR; 10.
DR Pfam; PF01463; LRRCT; 1.
DR Pfam; PF01462; LRRNT; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00408; IGc2; 1.
DR SMART; SM00370; LRR; 6.
DR SMART; SM00082; LRRCT; 1.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_TYP; 2.

```
KW   Immunoglobulin domain; Transmembrane; Glycoprotein; Repeat;
KW   Leucine-rich repeat; Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    713       GLIOMA AMPLIFIED ON CHROMOSOME 1 PROTEIN.
FT   DOMAIN       19    630       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    631    651       POTENTIAL.
FT   DOMAIN      652    713       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       92    115       LRR 1.
FT   REPEAT      116    139       LRR 2.
FT   REPEAT      140    163       LRR 3.
FT   REPEAT      165    187       LRR 4.
FT   REPEAT      188    211       LRR 5.
FT   REPEAT      213    235       LRR 6.
FT   REPEAT      236    259       LRR 7.
FT   REPEAT      261    283       LRR 8.
FT   REPEAT      309    333       LRR 9.
FT   REPEAT      334    357       LRR 10.
FT   REPEAT      359    385       LRR 11.
FT   DOMAIN      438    504       IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID    445    497       BY SIMILARITY.
FT   CARBOHYD     94     94       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    381    381       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    555    555       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    583    583       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   713 AA;  78798 MW;  EC8BC0FD240C9396 CRC64;
```

```
  Query Match           13.0%;  Score 438.5;  DB 1;  Length 713;
  Best Local Similarity  25.2%;  Pred. No. 2.3e-22;
  Matches  147;  Conservative   86;  Mismatches  228;  Indels  123;  Gaps   12;

Qy      20 RALFDPLLVVLLALQLLVVAGLVRAQTCPSVCSCS-----------NQFSKVICVRKNLR 68
           | |  |||:  :|     |  :     ||  |:|             : : | |    |
Db       2 RLLVAPLLLAWVAGATAAVPVVPWHVPCPPQCACQIRPWYTPRSSYREATTVDCNDLFLT 61

Qy      69 EVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNT 128
            ||  :   |: | |  | |  :  :   :| :|  | ||:|          |: |  | :
Db      62 AVPPALPAGTQTLLLQSNSIVRVDQSELGYLANLTELDLSQNSFSDARDCDFHALPQLLS 121

Qy     129 LELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGE------- 181
           | | :|:|| : : :|  |: |:||:| :| :  |   ||: : :| || |
Db     122 LHLEENQLTRLEDHSFAGLASLQELYLNHNQLYRIAPRAFSGLSNLLRLHLNSNLLRAID 181

Qy     182 ------LKRLSYISEGA----------FEGLSNLRYLNLAMCNLREIPN----------- 214
                |  |  :  |           |  |:||| | ||  ||||| :
Db     182 SRWFEMLPNLEILMIGGNKVDAILDMNFRPLANLRSLVLAGMNLREISDYALEGLQSLES 241

Qy     215 --------------LTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQ----- 254
                         |  :  |  |||: | |  : || |  ::||::| :   :
Db     242 LSFYDNQLARVPRRALEQVPGLKFLDLNKNPLQRVGPGDFANMLHLKELGLNNMEELVSI 301

Qy     255 ---------------------IQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHL 293
                             :  |   || :|  :  : | :| |: |      | :|
Db     302 DKFALVNLPELTKLDITNNPRLSFIHPRAFHHLPQMETLMLNNNALSALHQQTVESLPNL 361

Qy     294 ERIHLHHNPWNCNCDILWLS---WWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFT 350
           : : || ||  |:| | | :     :: : | :| |      ||:|:   : |:     |
Db     362 QEVGLHGNPIRCDCVIRWANATGTRVRFIEPQSTLCA----EPPDLQRLPVREVPFREMT 417

Qy     351 CYAPVIVEP---PADLNVTEGMAAELKCRA-STSLTSVSWITPNGTVMTHGAYKVRIAVL 406
           :   :: |   |  | |  | :  | ||| :     : |:|| |  :|      |  |
Db     418 DHCLPLISPRSFPPSLQVASGESMVLHCRALAEPEPEIYWVTPAGLRLTPAHAGRRCRVY 477

Qy     407 SDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQ 466
           :|||    || :: |:|||:  | ||  | : :: |  |              ::| :
Db     478 PEGTLELRRVTAEEAGLYTCVAQNLVGADTKTVSVVVGRA-------------LLQPGR 523

Qy     467 DEARTTDNNVGPTP----VVDWETTNVTTSLTPQSTRSTEKTFT 506
           || :  :  |  |     :: |        :|| :| ||  |::
Db     524 DEGQGLELRVQETHPYHILLSW--------VTPPNTVSTNLTWS 559


RESULT   2
```

```
CHAD_BOVIN
ID   CHAD_BOVIN     STANDARD;      PRT;   361 AA.
AC   Q27972;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Chondroadherin precursor (Cartilage leucine-rich protein) (38 kDa bone
DE   protein).
GN   CHAD.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Cartilage;
RX   MEDLINE=94342341; PubMed=8063792;
RA   Neame P.J., Sommarin Y., Boynton R.E., Heinegaard D.;
RT   "The structure of a 38-kDa leucine-rich protein (chondroadherin)
RT   isolated from bovine cartilage.";
RL   J. Biol. Chem. 269:21547-21554(1994).
RN   [2]
RP   SEQUENCE OF 25-55 AND 77-97.
RC   TISSUE=Bone;
RX   MEDLINE=95113864; PubMed=7814406;
RA   Hu B., Coulson L., Moyer B., Price P.A.;
RT   "Isolation and molecular cloning of a novel bone phosphoprotein
RT   related in sequence to the cystatin family of thiol protease
RT   inhibitors.";
RL   J. Biol. Chem. 270:431-436(1995).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U08018; AAA21330.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 10.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 5.
KW   Repeat; Signal.
FT   SIGNAL        1     24       OR 23 (IN SOME ISOFORM(S)).
FT   CHAIN        25    361       CHONDROADHERIN.
FT   CHAIN        25    352       CHONDROADHERIN, MINOR FORM.
FT   DOMAIN       79    317       10 X 24 AA LEUCINE-RICH TANDEM REPEATS.
FT   REPEAT       79    102       1.
FT   REPEAT      103    126       2.
FT   REPEAT      127    150       3.
FT   REPEAT      151    174       4.
FT   REPEAT      175    198       5.
FT   REPEAT      199    222       6.
FT   REPEAT      223    246       7.
FT   REPEAT      248    271       8.
FT   REPEAT      272    293       9.
FT   REPEAT      294    317       10.
FT   DISULFID    306    348
FT   DISULFID    308    328
FT   CONFLICT     25     25       C -> Y (IN REF. 2).
FT   CONFLICT     29     29       C -> W (IN REF. 2).
FT   CONFLICT     31     31       C -> H (IN REF. 2).
FT   CONFLICT     40     40       C -> L (IN REF. 2).
FT   CONFLICT     52     52       S -> R (IN REF. 2).
```

```
SQ   SEQUENCE   361 AA;  40884 MW;  DA79DC98AD3DD1F8 CRC64;


  Query Match             12.2%;  Score 411.5;  DB 1;  Length 361;
  Best Local Similarity   32.4%;  Pred. No. 6.3e-21;
  Matches  113;  Conservative   54;  Mismatches  145;  Indels   37;  Gaps   10;

Qy      25 PLLVVLLALQLLVVAGLVRA-QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNL 83
           |:|:: |:| ||  | |: |   ||  | | :    ||| :  |:::|  :|  |:|||
Db       4 PMLLLXLSLGLL--ASLLPALAACPQNCHCHSDLQHVICDKVGLQKIPK-VSEKTKLLNL 60

Qy      84 HENQIQIIKVNSFK--------HLRHLEI----------------LQLSRNHIRTIEIGA 119
             |   ::  |||:         ||:| :|                | || | || :  ||
Db      61 QRNNFPVLATNSFRAMPNLVSLHLQHCQIREVAAGAFRGLKQLIYLYLSHNDIRVLRAGA 120

Qy     120 FNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDL 179
           |: |  |  | |  |::| :| |     |  |  | | || |  : | ||      || | |
Db     121 FDDLTELTYLYLDHNKVTELPRGLLSPLVNLFILQLNNNKIRELRSGAFQGAKDLRWLYL 180

Qy     180 GELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPN--LTPLIKLDELDLSGNHLSAIRPG 237
             |   || :  || : : ||    |    |   |:  |: |  ::|| || | | :|
Db     181 SE-NSLSSLQPGALDDVENLAKFYLDRNQLSSYPSAALSKLRVVEELKLSHNPLKSIPDN 239

Qy     238 SFQGL-MHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERI 296
           :||    :|: ||:  : ::      ||  : :|  ::| :| |  || :   |   || :
Db     240 AFQSFGRYLETLWLDNTNLEKFSDGAFLGVTTLKHVHLENNRLHQLPSNF--PFDSLETL 297

Qy     297 HLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELD 345
           | :||| | | :  |  |::       | ||   :|   :|::| : |
Db     298 TLTNNPWKCTCQLRGLRRWLEAKTSRPDATCA---SPAKFRGQHIRDTD 343


RESULT   3
GPV_MOUSE
ID   GPV_MOUSE      STANDARD;      PRT;   567 AA.
AC   O08742;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Platelet glycoprotein V precursor (GPV) (CD42D).
GN   GP5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Liver;
RX   MEDLINE=97275136; PubMed=9129030;
RA   Ravanat C., Morales M., Azorsa D.O., Moog S., Schuhler S.,
RA   Grunert P., Loew D., van Dorsselaer A., Cazenave J.-P., Lanza F.;
RT   "Gene cloning of rat and mouse platelet glycoprotein V:
RT   identification of megakaryocyte-specific promoters and demonstration
RT   of functional thrombin cleavage.";
RL   Blood 89:3253-3262(1997).
CC   -!- FUNCTION: THE GPIB-V-IX COMPLEX FUNCTIONS AS THE VON WILLEBRAND
CC       FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT
CC       PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO
CC       INJURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A
CC       CRITICAL INITIATING EVENT IN HEMOSTASIS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: CONTAINS 14 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z69595; CAA93441.1; -.
DR   MGD; MGI:1096363; Gp5.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 13.
DR   Pfam; PF01463; LRRCT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00370; LRR; 1.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 10.
KW   Platelet; Transmembrane; Glycoprotein; Blood coagulation;
KW   Repeat; Leucine-rich repeat; Cell adhesion; Signal.
FT   SIGNAL       1     16       POTENTIAL.
FT   CHAIN       17    567       PLATELET GLYCOPROTEIN V.
FT   DOMAIN      17    522       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   523    543       POTENTIAL.
FT   DOMAIN     544    567       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      73     96       LRR 1.
FT   REPEAT      97    120       LRR 2.
FT   REPEAT     122    144       LRR 3.
FT   REPEAT     145    168       LRR 4.
FT   REPEAT     170    192       LRR 5.
FT   REPEAT     194    216       LRR 6.
FT   REPEAT     217    240       LRR 7.
FT   REPEAT     241    264       LRR 8.
FT   REPEAT     266    288       LRR 9.
FT   REPEAT     289    312       LRR 10.
FT   REPEAT     314    337       LRR 11.
FT   REPEAT     338    361       LRR 12.
FT   REPEAT     362    385       LRR 13.
FT   REPEAT     386    409       LRR 14.
FT   CARBOHYD    51     51       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    67     67       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   243    243       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   298    298       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   312    312       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   385    385       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   567 AA;  63467 MW;  C48643AA73967A7D CRC64;


  Query Match             11.0%;  Score 369.5;  DB 1;  Length 567;
  Best Local Similarity   30.1%;  Pred. No. 8e-18;
  Matches  119;  Conservative   63;  Mismatches  166;  Indels   47;  Gaps   12;

Qy       1 MLNKMTLHPQQIMIGPRFNRALFDPLLVVLLALQLLVVAGLVRAQTC---PSVCSCSNQF 57
             :|:|| |  :|: :     | |   |   |  || |   || : |      :: |   :
Db     117 ILDKMVL-LEQLFLDHNALRDLDQNLFQQLRNLQEL---GLNQNQLSFLPANLFSSLREL 172

Qy      58 SKVICVRKNLREVPDGI---STNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRT 114
             :   | ||  :| |:         | |: ||:  :      :|  |  |:| |||:|:
Db     173 KLLDLSRNNLTHLPKGLLGAQVKLEKLLLYSNQLTSVDSGLLSNLGALTELRLERNHLRS 232

Qy     115 IEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSL 174
           :  |||: | ||::| |  | | ::|   |:::| :  | |  ||:| :|   |  :  |
Db     233 VAPGAFDRLGNLSSLTLSGNLLESLPPALFLHVSSVSRLTLFENPLEELPDVLFGEMAGL 292

Qy     175 RRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIK-----LDELDLSGN 229
           | | |     ||  :   ||  || |: | |        | |: | :     | ||  : |
Db     293 RELWLNG-THLSTLPAAAFRNLSGLQTLGLT-----RNPRLSALPRGVFQGLRELRVLGL 346

Qy     230 H---LSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDL 286
           |   |: :|  : :|| ||::: :  :::: : |  | || ||  : | ||  |  || |:
Db     347 HTNALAELRDDALRGLGHLRQVSLRHNRLRALPRTLFRNLSSLESVQLEHNQLETLPGDV 406

Qy     287 FTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQ 346
           |  |  | :: | |||| |:| :     |::                 |:: ||   |  |
Db     407 FAALPQLTQVLLGHNPWLCDCGLWRFLQWLRH-------------HPDILGR--DEPPQ 450

Qy     347 NYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSL 381
                 |  |   || | |:  | :  :  |    ||
```

```
Db    451 ----CRGP---EPRASLSFWELLQGDPWCPDPRSL 478


RESULT   4
GPV_RAT
ID   GPV_RAT        STANDARD;      PRT;   567 AA.
AC   O08770;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Platelet glycoprotein V precursor (GPV) (CD42D).
GN   GP5.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR; TISSUE=Liver;
RX   MEDLINE=97275136; PubMed=9129030;
RA   Ravanat C., Morales M., Azorsa D.O., Moog S., Schuhler S.,
RA   Grunert P., Loew D., van Dorsselaer A., Cazenave J.-P., Lanza F.;
RT   "Gene cloning of rat and mouse platelet glycoprotein V:
RT   identification of megakaryocyte-specific promoters and demonstration
RT   of functional thrombin cleavage.";
RL   Blood 89:3253-3262(1997).
CC   -!- FUNCTION: THE GPIB-V-IX COMPLEX FUNCTIONS AS THE VON WILLEBRAND
CC       FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT
CC       PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO
CC       INJURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A
CC       CRITICAL INITIATING EVENT IN HEMOSTASIS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: CONTAINS 14 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z69594; CAA93440.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 14.
DR   Pfam; PF01463; LRRCT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00370; LRR; 2.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 10.
KW   Platelet; Transmembrane; Glycoprotein; Blood coagulation;
KW   Repeat; Leucine-rich repeat; Cell adhesion; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    567       PLATELET GLYCOPROTEIN V.
FT   DOMAIN       17    522       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    523    543       POTENTIAL.
FT   DOMAIN      544    567       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       73     96       LRR 1.
FT   REPEAT       97    120       LRR 2.
FT   REPEAT      122    144       LRR 3.
FT   REPEAT      145    168       LRR 4.
FT   REPEAT      169    192       LRR 5.
FT   REPEAT      194    216       LRR 6.
FT   REPEAT      217    240       LRR 7.
FT   REPEAT      241    264       LRR 8.
FT   REPEAT      266    288       LRR 9.
FT   REPEAT      289    312       LRR 10.
FT   REPEAT      314    337       LRR 11.
FT   REPEAT      338    361       LRR 12.
FT   REPEAT      362    385       LRR 13.
FT   REPEAT      387    409       LRR 14.
FT   CARBOHYD     51     51       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    243    243       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    298    298       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    312    312       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    385    385       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    498    498       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   567 AA;  63344 MW;  CA10708E0D03707F CRC64;


  Query Match             10.8%;  Score 361.5;  DB 1;  Length 567;
  Best Local Similarity   29.0%;  Pred. No. 2.8e-17;
  Matches  121;  Conservative   59;  Mismatches  180;  Indels   57;  Gaps   11;

Qy      1 MLNKMTLHPQQIMIGPRFNRALFDPLLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFS-- 58
          :|:|| |  :|: :    | |   |   || |:              :|   || |
Db    117 ILDKMVL-LEQLFLDHNALRDLDQNLFQKLLNLR-------------DLCLNQNQLSFL 161

Qy     59 ----------KVICV-RKNLREVPDGI---STNTRLLNLHENQIQIIKVNSFKHLRHLE 103
                    ||: : | ||  :| |:         | |: |::  :      :| |
Db    162 PANLFSSLGKLKVLDLSRNNLTHLPQGLLGAQIKLEKLLLYSNRLMSLDSGLLANLGALT 221

Qy    104 ILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESI 163
          |:| |||:|:|  |||: | ||:|| |  | | ::|   |:::| |  | |  ||:| :
Db    222 ELRLERNHLRSIAPGAFDSLGNLSTLTLSGNLLESLPPALFLHVSWLTRLTLFENPLEEL 281

Qy    164 PSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTP-----L 218
          |   |  :  || | |     |  :   ||  || |: | |    |  :  | |     |
Db    282 PEVLFGEMAGLRELWLNG-THLRTLPAAAFRNLSGLQTLGLTRNPL--LSALPPGMFHGL 338

Qy    219 IKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNN 278
          :|  | :  | |  :   : :||  |::: :  :::: : |  | || ||| : | ||
Db    339 TELRVLAVHTNALEELPEDALRGLGRLRQVSLRHNRLRALPRTLFRNLSSLVTVQLEHNQ 398

Qy    279 LTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWI--------KDMAPSNTACCARC 330
          |  || |:|  |  | |: | |||| |:| :     |:        :|  |       :|
Db    399 LKTLPGDVFAALPQLTRVLLGHNPWLCDCGLWPFLQWLRHHLELLGRDEPP-------QC 451

Qy    331 NTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWI 387
          | | :      || |   |  |     | |     : |   :  |  | :|:
Db    452 NGPESRASLTFWELLQGDQWC--PSSRGLPPDPPTENALKAPDPTQRPNSSQSWAWV 506


RESULT   5
SLIT_DROME
ID   SLIT_DROME     STANDARD;      PRT;  1480 AA.
AC   P24014;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Slit protein precursor.
GN   SLI.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91099665; PubMed=2176636;
RA   Rothberg J.M., Jacobs J.R., Goodman C.S., Artavanis-Tsakonas S.;
RT   "Slit: an extracellular protein necessary for development of midline
RT   glia and commissural axon pathways contains both EGF and LRR
RT   domains.";
RL   Genes Dev. 4:2169-2187(1990).
CC   -!- FUNCTION: NECESSARY FOR DEVELOPMENT OF MIDLINE GLIA AND
CC       COMMISSURAL AXON PATHWAYS. SLIT MAY INTERACT WITH EXTRACELLULAR
CC       MATRIX MOLECULES.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
```

```
CC   -!- TISSUE SPECIFICITY: EXCRETED BY THE MIDLINE GLIA CELLS AND
CC       EVENTUALLY DISTRIBUTED ALONG THE AXONS.
CC   -!- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 24 LEUCINE-RICH REPEATS (LRR).
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   -!- SIMILARITY: CONTAINS 1 LAMININ G-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X53959; CAA37910.1; -.
DR   PIR; A36665; A36665.
DR   HSSP; P00743; 1CCF.
DR   FlyBase; FBgn0003425; sli.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000359; Cys_knot.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00007; Cys_knot; 1.
DR   Pfam; PF00008; EGF; 7.
DR   Pfam; PF00054; laminin_G; 1.
DR   Pfam; PF00560; LRR; 17.
DR   Pfam; PF01463; LRRCT; 4.
DR   Pfam; PF01462; LRRNT; 4.
DR   SMART; SM00041; CT; 1.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00001; EGF_like; 5.
DR   SMART; SM00370; LRR; 4.
DR   SMART; SM00082; LRRCT; 4.
DR   SMART; SM00013; LRRNT; 4.
DR   SMART; SM00369; LRR_TYP; 9.
DR   SMART; SM00282; LamG; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS00022; EGF_1; 7.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
KW   Neurogenesis; Glycoprotein; Signal; Alternative splicing;
KW   EGF-like domain; Repeat; Leucine-rich repeat.
FT   SIGNAL       1     36
FT   CHAIN       37   1480       SLIT PROTEIN.
FT   REPEAT      99    122       LRR 1.
FT   REPEAT     123    146       LRR 2.
FT   REPEAT     148    170       LRR 3.
FT   REPEAT     171    194       LRR 4.
FT   REPEAT     195    218       LRR 5.
FT   REPEAT     220    246       LRR 6.
FT   REPEAT     321    344       LRR 7.
FT   REPEAT     345    368       LRR 8.
FT   REPEAT     369    392       LRR 9.
FT   REPEAT     394    416       LRR 10.
FT   REPEAT     417    440       LRR 11.
FT   REPEAT     522    545       LRR 12.
FT   REPEAT     546    569       LRR 13.
FT   REPEAT     570    593       LRR 14.
FT   REPEAT     595    617       LRR 15.
FT   REPEAT     618    641       LRR 16.
FT   REPEAT     643    666       LRR 17.
FT   REPEAT     678    701       LRR 18.
FT   REPEAT     720    743       LRR 19.
FT   REPEAT     745    764       LRR 20.
FT   REPEAT     765    788       LRR 21.
FT   REPEAT     790    812       LRR 22.
FT   REPEAT     813    836       LRR 23.
FT   REPEAT     838    861       LRR 24.
FT   DOMAIN     907    944       EGF-LIKE 1.
FT   DOMAIN     946    983       EGF-LIKE 2.
FT   DOMAIN     985   1022       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1024   1062       EGF-LIKE 4.
FT   DOMAIN    1064   1100       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1111   1149       EGF-LIKE 6.
FT   DOMAIN    1152   1325       LAMININ G-LIKE.
FT   DOMAIN    1353   1392       EGF-LIKE 7.
FT   DOMAIN    1409   1480       CTCK.
FT   CARBOHYD   111    111       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   207    207       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   357    357       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   435    435       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   783    783       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   788    788       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   958    958       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   998    998       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1060   1060       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1159   1159       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1175   1175       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1243   1243       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1292   1292       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID   911    922       BY SIMILARITY.
FT   DISULFID   916    932       BY SIMILARITY.
FT   DISULFID   934    943       BY SIMILARITY.
FT   DISULFID   950    961       BY SIMILARITY.
FT   DISULFID   955    971       BY SIMILARITY.
FT   DISULFID   973    982       BY SIMILARITY.
FT   DISULFID   989   1001       BY SIMILARITY.
FT   DISULFID   995   1010       BY SIMILARITY.
FT   DISULFID  1012   1021       BY SIMILARITY.
FT   DISULFID  1028   1041       BY SIMILARITY.
FT   DISULFID  1035   1050       BY SIMILARITY.
FT   DISULFID  1052   1061       BY SIMILARITY.
FT   DISULFID  1068   1079       BY SIMILARITY.
FT   DISULFID  1073   1088       BY SIMILARITY.
FT   DISULFID  1090   1099       BY SIMILARITY.
FT   DISULFID  1115   1125       BY SIMILARITY.
FT   DISULFID  1120   1137       BY SIMILARITY.
FT   DISULFID  1139   1148       BY SIMILARITY.
FT   DISULFID  1357   1368       BY SIMILARITY.
FT   DISULFID  1362   1380       BY SIMILARITY.
FT   DISULFID  1382   1391       BY SIMILARITY.
FT   DISULFID  1409   1443       BY SIMILARITY.
FT   DISULFID  1423   1457       BY SIMILARITY.
FT   DISULFID  1434   1473       BY SIMILARITY.
FT   DISULFID  1438   1475       BY SIMILARITY.
FT   DISULFID  1442   1479       BY SIMILARITY.
FT   VARSPLIC  1394   1404       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   1480 AA;  165752 MW;  F9D5925FC170B1C3 CRC64;


  Query Match             10.4%;  Score 350;  DB 1;  Length 1480;
  Best Local Similarity   23.9%;  Pred. No. 5.8e-16;
  Matches  109;  Conservative   65;  Mismatches  152;  Indels  130;  Gaps    9;

Qy      46 TCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEIL 105
           :||  | |::    | |  |:|  ||  :  :|  : | :| |  :   ||   | |  :
Db     294 SCPHPCRCADGI--VDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKSFSSFRRLRRI 351

Qy     106 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 165
           || |:|  |   | :||  | || |: |::  :|:| |  |  |: | |  | |  |
Db     352 DLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLLNANEISCIRK 411

Qy     166 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM------CNLR--------- 210
           ||  : ||  | | :   :  :: | |: : ::: ::||       ||||
Db     412 DAFRDLHSLSLLSLYD-NNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRWLADYLHKN 470
```

```
Qy    211 ------------------------------------------------------------ 210

Db    471 PIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKLSGECRMDSDCPAMCHCEGTTV 530

Qy    211 --------EIPNLTPL--------------IKLD----------ELDLSGNHLSAIRPGS 238
                  |||   ||             |  |           :|:|  | |: | | :
Db    531 DCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKRNQLTGIEPNA 590

Qy    239 FQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 298
          |:|  |:|:| : :::|: |     |  |  |  :||  | :: :     |  |: |  ::|
Db    591 FEGASHIQELQLGENKIKEISNKMFLGLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNL 650

Qy    299 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVE 358
            ||:|||| : | :  ::  : :  |  ||| |  ::   | :|  : | | :
Db    651 ASNPFNCNCHLAWFAECVRKKSLNGGA--ARCGAPSKVRDVQIKDLPHSEFKCSS----- 703

Qy    359 PPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVM 394
                :||   :  |  | : |        |||:
Db    704 -----ENSEGCLGDGYCPPSCTCT--------GTVV 726


RESULT   6
ALS_PAPHA
ID   ALS_PAPHA      STANDARD;      PRT;   605 AA.
AC   O02833;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Insulin-like growth factor binding protein complex acid labile chain
DE   precursor (ALS).
GN   IGFALS OR ALS.
OS   Papio hamadryas (Hamadryas baboon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Papio.
OX   NCBI_TaxID=9557;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=97040714; PubMed=8886027;
RA   Delhanty P., Baxter R.C.;
RT   "The cloning and expression of the baboon acid-labile subunit of the
RT   insulin-like growth factor binding protein complex.";
RL   Biochem. Biophys. Res. Commun. 227:897-902(1996).
CC   -!- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS THAT RESULT
CC       IN PROTEIN COMPLEXES, RECEPTOR-LIGAND BINDING OR CELL ADHESION.
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 kDa WITH
CC       IGF-I OR IGF-II AND IGFBP-3 (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- SIMILARITY: CONTAINS 20 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S83462; -; NOT_ANNOTATED_CDS.
DR   HSSP; P23945; 1XUN.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 19.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00370; LRR; 2.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 11.
KW   Glycoprotein; Leucine-rich repeat; Repeat; Signal.
FT   SIGNAL        1     27       BY SIMILARITY.
FT   CHAIN        28    605       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                PROTEIN COMPLEX ACID LABILE CHAIN.
FT   REPEAT       53     73       LRR 1.
FT   REPEAT       74     96       LRR 2.
FT   REPEAT       98    120       LRR 3.
FT   REPEAT      121    144       LRR 4.
FT   REPEAT      146    168       LRR 5.
FT   REPEAT      169    192       LRR 6.
FT   REPEAT      194    216       LRR 7.
FT   REPEAT      217    240       LRR 8.
FT   REPEAT      242    264       LRR 9.
FT   REPEAT      265    288       LRR 10.
FT   REPEAT      289    312       LRR 11.
FT   REPEAT      313    336       LRR 12.
FT   REPEAT      338    360       LRR 13.
FT   REPEAT      361    384       LRR 14.
FT   REPEAT      386    408       LRR 15.
FT   REPEAT      409    432       LRR 16.
FT   REPEAT      433    456       LRR 17.
FT   REPEAT      458    480       LRR 18.
FT   REPEAT      482    504       LRR 19.
FT   REPEAT      505    530       LRR 20.
FT   CARBOHYD     64     64       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     85     85       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     96     96       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    368    368       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    515    515       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    580    580       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   605 AA;  66110 MW;  9D71994625F23652 CRC64;


  Query Match             10.0%;  Score 335;  DB 1;  Length 605;
  Best Local Similarity   20.8%;  Pred. No. 1.9e-15;
  Matches  125;  Conservative   64;  Mismatches  125;  Indels  286;  Gaps   13;

Qy     31 LALQLLVVA----------------GLVRAQTCPSVCSCS-----NQFSKVICVRKNLR 68
          ||| ||:::                |      ||: |:||     |: | | |  :||
Db      8 LALALLLLSWVALGPRSLEGAEPGTPGEAEGPACPATCACSYDDEVNELS-VFCSSRNLT 66

Qy     69 EVPDGISTNTRL---------------------------------------------- 80
            :||||   |:
Db     67 RLPDGIPGGTQALWLDSNNLSSIPPAAFRNLSSLAFLNLQGGQLGSLEPQALLGLENLCH 126

Qy     81 LNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIP 140
          |:|  ||:: : | :| :   | :| || | :  :| | | || ||  | |  | |  :|
Db    127 LHLERNQLRSLAVGTFAYTPALALLGLSNNRLSRLEDGLFEGLGNLWDLNLGWNSLAVLP 186

Qy    141 NGAF----------------VY--------LSKLKELWLRNNPIESIPSYAFNRIPSLRR 176
          : ||                |        |::|:|| |  | : :| :  | ::| |::
Db    187 DAAFRGLGGLRELVLAGNRLAYLQPALFSGLAELRELDLSRNALRAIKANVFAQLPRLQK 246

Qy    177 L---------------------------------------------------------- 177
          |
Db    247 LYLDRNLIAAVAPGAFLGLKALRWLDLSHNRVAGLLEDTFPGLLGLRVLRLSHNAIASLR 306

Qy    178 ----------------------------DLGELK-------RLSYISEGAFEGLSNLRY 201
                                       ||:|:       :|  :  ||| ||:|:
Db    307 PRTFEDLHFLEELQLGHNRIRQLAERSFEGLGQLEVLTLDHNQLQEVKVGAFLGLTNVAV 366

Qy    202 LNLAMCNLREIPN--LTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIE 259
          :||:   || :|      | ||  | | |: |  ||| :| ||  |::|::  :  :  ||
Db    367 MNLSGNCLRNLPEQVFRGLGKLHSLHLEGSCLGRIRPHTFAGLSGLRRLFLKDNGLVGIE 426

Qy    260 RNAFDNLQSLVEINLAHNNLTLLPHDLFTPL---------------------------- 290
            :   |  |:|::|  | || ||| ||  |
Db    427 EQSLWGLAELLELDLTSNQLTHLPHQLFQGLGKLEYLLLSHNRLAELPADALGPLQRAFW 486

Qy    291 ----HH-----------------------------------LERIHLHHNPWNCNCDILW 311
              |:                                   |||: |  |||:|:| :
```

```
Db    487 LDVSHNRLEALPGSLLASLGRLRYLNLRNNSLRTFTPQPPGLERLWLEGNPWDCSCPLKA 546

Qy    312 LSWWIKDMAPSNTACCAR----------------------CNTPPNLKGRYIGELDQNYF 349
          |    :| |  | :   |                      | :|| : |  : :| : :|
Db    547 L----RDFALQNPSAVPRFVQAICEGDDCQPPVYTYNNITCASPPEVAGLDLRDLGEAHF 602


RESULT   7
GPV_HUMAN
ID   GPV_HUMAN      STANDARD;      PRT;   560 AA.
AC   P40197;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Platelet glycoprotein V precursor (GPV) (CD42D).
GN   GP5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=93391348; PubMed=7690959;
RA   Hickey M.J., Hagen F.S., Yagi M., Roth G.J.;
RT   "Human platelet glycoprotein V: characterization of the polypeptide
RT   and the related Ib-V-IX receptor system of adhesive, leucine-rich
RT   glycoproteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:8327-8331(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Platelet;
RX   MEDLINE=94012616; PubMed=8407908;
RA   Lanza F., Morales M., de la Salle C., Cazenave J.-P., Clemetson K.J.,
RA   Shimomura T., Phillips D.R.;
RT   "Cloning and characterization of the gene encoding the human platelet
RT   glycoprotein V. A member of the leucine-rich glycoprotein family
RT   cleaved during thrombin-induced platelet activation.";
RL   J. Biol. Chem. 268:20801-20807(1993).
RN   [3]
RP   PARTIAL SEQUENCE.
RC   TISSUE=Platelet;
RX   MEDLINE=90275263; PubMed=2350580;
RA   Shimomura T., Fujimura K., Maehama S., Takemoto M., Oda K.,
RA   Fujimoto T., Oyama R., Suzuki M., Icihara-Tanaka K., Titani K.,
RA   Kuramoto A.;
RT   "Rapid purification and characterization of human platelet
RT   glycoprotein V: the amino acid sequence contains leucine-rich
RT   repetitive modules as in glycoprotein Ib.";
RL   Blood 75:2349-2356(1990).
RN   [4]
RP   PARTIAL SEQUENCE.
RC   TISSUE=Platelet;
RX   MEDLINE=90321220; PubMed=2372284;
RA   Roth G.J., Church T.A., McMullen B.A., Williams S.A.;
RT   "Human platelet glycoprotein V: a surface leucine-rich glycoprotein
RT   related to adhesion.";
RL   Biochem. Biophys. Res. Commun. 170:153-161(1990).
CC   -!- FUNCTION: THE GPIB-V-IX COMPLEX FUNCTIONS AS THE VON WILLEBRAND
CC       FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT
CC       PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO
CC       INJURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A
CC       CRITICAL INITIATING EVENT IN HEMOSTASIS.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: PLATELETS AND MEGAKARYOCYTES.
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   -!- SIMILARITY: CONTAINS 14 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L11238; AAA03069.1; -.
DR   EMBL; Z23091; CAA80637.1; -.
DR   HSSP; P16473; 1XUM.
DR   MIM; 173511; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 14.
DR   Pfam; PF01463; LRRCT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00370; LRR; 1.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 9.
KW   Platelet; Transmembrane; Glycoprotein; Blood coagulation;
KW   Repeat; Leucine-rich repeat; Cell adhesion; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    560       PLATELET GLYCOPROTEIN V.
FT   DOMAIN       17    523       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    524    544       POTENTIAL.
FT   DOMAIN      545    560       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       73     96       LRR 1.
FT   REPEAT       97    120       LRR 2.
FT   REPEAT      122    144       LRR 3.
FT   REPEAT      145    168       LRR 4.
FT   REPEAT      169    192       LRR 5.
FT   REPEAT      194    216       LRR 6.
FT   REPEAT      217    240       LRR 7.
FT   REPEAT      241    264       LRR 8.
FT   REPEAT      266    288       LRR 9.
FT   REPEAT      289    312       LRR 10.
FT   REPEAT      314    337       LRR 11.
FT   REPEAT      338    361       LRR 12.
FT   REPEAT      362    385       LRR 13.
FT   REPEAT      386    409       LRR 14.
FT   CARBOHYD     51     51       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    243    243       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    267    267       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    298    298       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    312    312       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    385    385       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    499    499       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     73     74       MT -> TK (IN REF. 2).
FT   CONFLICT    109    109       K -> T (IN REF. 2).
FT   CONFLICT    130    130       D -> W (IN REF. 3).
FT   CONFLICT    136    138       GID -> PGG (IN REF. 3).
FT   CONFLICT    209    209       L -> I (IN REF. 2).
FT   CONFLICT    267    267       N -> H (IN REF. 3).
FT   CONFLICT    327    327       L -> I (IN REF. 2).
FT   CONFLICT    478    478       P -> G (IN REF. 2).
FT   CONFLICT    509    509       P -> D (IN REF. 2).
SQ   SEQUENCE   560 AA;  60959 MW;  B1CDB04AF8AF7115 CRC64;


  Query Match             9.8%;  Score 329.5;  DB 1;  Length 560;
  Best Local Similarity   29.7%;  Pred. No. 4e-15;
  Matches   83;  Conservative   43;  Mismatches  104;  Indels   49;  Gaps    3;

Qy     66 NLREVPDG-----------------ISTNTRLLN---------LHENQIQIIKVNSFKH 98
          ||  :| |                  :| :: |||         | | |: |   :|
Db    181 NLTHLPKGLLGAQAKLERLLLHSNRLVSLDSGLLNSLGALTELQFHRNHIRSIAPGAFDR 240

Qy     99 LRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNN 158
          | :|  | |||||:  :    |   ||  | ||:| |  :|  |  :  |:||||
Db    241 LPNLSSLTLSRNHLAFLPSALFLHSHNLTLLTLFENPLAELPGVLFGEMGGLQELWLNRT 300

Qy    159 PIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPL 218
```

```
          : ::|: ||  :  || | :    ||| : :|||:||  |:
Db    301 QLRTLPAAAFRNLSRLRYLGVTLSPRLSALPQGAFQGLGELQV---------------- 343

Qy    219 IKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNN 278
              | |  | |:|:  |  :||  |::: : ::::: : |  | || ||  : | ||
Db    344 -----LALHSNGLTALPDGLLRGLGKLRQVSLRRNRLRALPRALFRNLSSLESVQLDHNQ 398

Qy    279 LTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIK 317
          |  || |:|  |  |  : | || | |:| :     |::
Db    399 LETLPGDVFGALPRLTEVLLGHNSWRCDCGLGPFLGWLR 437


RESULT   8
ALS_HUMAN
ID   ALS_HUMAN      STANDARD;      PRT;   605 AA.
AC   P35858;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Insulin-like growth factor binding protein complex acid labile chain
DE   precursor (ALS).
GN   IGFALS OR ALS.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Liver;
RX   MEDLINE=92357025; PubMed=1379671;
RA   Leong S.R., Baxter R.C., Camerato T., Dai J., Wood W.I.;
RT   "Structure and functional expression of the acid-labile subunit of
RT   the insulin-like growth factor-binding protein complex.";
RL   Mol. Endocrinol. 6:870-876(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Frankland J.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 28-35.
RX   MEDLINE=89308584; PubMed=2473065;
RA   Baxter R.C., Martin J.L., Beniac V.A.;
RT   "High molecular weight insulin-like growth factor binding protein
RT   complex. Purification and properties of the acid-labile subunit from
RT   human serum.";
RL   J. Biol. Chem. 264:11843-11848(1989).
CC   -!- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS THAT RESULT
CC       IN PROTEIN COMPLEXES, RECEPTOR-LIGAND BINDING OR CELL ADHESION.
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 kDa WITH
CC       IGF-I OR IGF-II AND IGFBP-3.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: CONTAINS 20 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M86826; AAA36047.1; -.
DR   EMBL; AL031724; CAC36078.1; -.
DR   PIR; A41915; A41915.
DR   HSSP; P23945; 1XUN.
DR   MIM; 601489; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 19.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00370; LRR; 2.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 11.
KW   Glycoprotein; Leucine-rich repeat; Repeat; Signal.
FT   SIGNAL        1     27
FT   CHAIN        28    605       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                 PROTEIN COMPLEX ACID LABILE CHAIN.
FT   REPEAT        53     73       LRR 1.
FT   REPEAT        74     96       LRR 2.
FT   REPEAT        98    120       LRR 3.
FT   REPEAT       121    144       LRR 4.
FT   REPEAT       145    168       LRR 5.
FT   REPEAT       170    192       LRR 6.
FT   REPEAT       193    216       LRR 7.
FT   REPEAT       217    240       LRR 8.
FT   REPEAT       242    264       LRR 9.
FT   REPEAT       265    288       LRR 10.
FT   REPEAT       289    312       LRR 11.
FT   REPEAT       313    336       LRR 12.
FT   REPEAT       337    360       LRR 13.
FT   REPEAT       361    384       LRR 14.
FT   REPEAT       386    408       LRR 15.
FT   REPEAT       409    432       LRR 16.
FT   REPEAT       433    456       LRR 17.
FT   REPEAT       458    480       LRR 18.
FT   REPEAT       482    504       LRR 19.
FT   REPEAT       505    530       LRR 20.
FT   CARBOHYD      64     64       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      85     85       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      96     96       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     368    368       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     515    515       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     580    580       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   605 AA;  66034 MW;  F6562A23CBE918F6 CRC64;


  Query Match             9.5%;  Score 320;  DB 1;  Length 605;
  Best Local Similarity   29.8%;  Pred. No. 2e-14;
  Matches   95;  Conservative   50;  Mismatches  124;  Indels   50;  Gaps    6;

Qy     31 LALQLLVVA----------------GLVRAQTCPSVCSCS-----NQFSKVICVRKNLR 68
          ||| ||:::                |     ||: | ||     :: | | |  :||
Db      8 LALALLLLSWVALGPRSLEGADPGTPGEAEGPACPAACVCSYDDDADELS-VFCSSRNLT 66

Qy     69 EVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI---------------------- 104
          :|||:   |: | |  | :  :   :|::|  |
Db     67 RLPDGVPGGTQALWLDGNNLSSVPPAAFQNLSSLGFLNLQGGQLGSLEPQALLGLENLCH 126

Qy    105 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 164
          | | || :|:: :| |     | :| | :|||: : :| |  |  | :| |  | :  :|
Db    127 LHLERNQLRSLALGTFAHTPALASLGLSNNRLSRLEDGLFEGLGSLWDLNLGWNSLAVLP 186

Qy    165 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPN--LTPLIKLD 222
            ||  : ||| | |    ||:|:    | ||: || |:|:   || |       | :|
Db    187 DAAFRGLGSLRELVLAG-NRLAYLQPALFSGLAELRELDLSRNALRAIKANVFVQLPRLQ 245

Qy    223 ELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLL 282
          :| |  | ::|: ||:| ||  |: | :  :::  :  : |  |  |. : |:|| :  |
Db    246 KLYLDRNLIAAVAPGAFLGLKALRWLDLSHNRVAGLLEDTFPGLLGLRVLRLSHNAIASL 305

Qy    283 PHDLFTPLHHLERIHLHHN 301
            |  || || : | ||
Db    306 RPRTFKDLHFLEELQLGHN 324


RESULT   9
ALS_RAT
ID   ALS_RAT        STANDARD;      PRT;   603 AA.
AC   P35859;
```

```
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Insulin-like growth factor binding protein complex acid labile chain
DE   precursor (ALS).
GN   IGFALS OR ALS.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=93038676; PubMed=1384485;
RA   Dai J., Baxter R.C.;
RT   "Molecular cloning of the acid-labile subunit of the rat insulin-like
RT   growth factor binding protein complex.";
RL   Biochem. Biophys. Res. Commun. 188:304-309(1992).
RN   [2]
RP   SEQUENCE OF 24-44, AND CHARACTERIZATION.
RC   STRAIN=WISTAR; TISSUE=Serum;
RX   MEDLINE=94130835; PubMed=7507839;
RA   Baxter R.C., Dai J.;
RT   "Purification and characterization of the acid-labile subunit of rat
RT   serum insulin-like growth factor binding protein complex.";
RL   Endocrinology 134:848-852(1994).
CC   -!- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN REGULATING THE ACCESS OF
CC       CIRCULATING IGFS TO THE TISSUES.
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 kDa WITH
CC       IGF-I OR IGF-II AND IGFBP-3.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: BRAIN, KIDNEY, LUNG, HEART, SPLEEN, MUSCLE
CC       AND LIVER.
CC   -!- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S46785; AAB23770.2; -.
DR   PIR; JC1282; JC1282.
DR   HSSP; P23945; 1XUN.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 19.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00370; LRR; 5.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 9.
KW   Glycoprotein; Leucine-rich repeat; Repeat; Signal.
FT   SIGNAL       1     23
FT   CHAIN       24    603       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                PROTEIN COMPLEX ACID LABILE CHAIN.
FT   REPEAT       52     73       LRR 1.
FT   REPEAT       74     96       LRR 2.
FT   REPEAT       98    120       LRR 3.
FT   REPEAT      121    144       LRR 4.
FT   REPEAT      146    168       LRR 5.
FT   REPEAT      169    192       LRR 6.
FT   REPEAT      194    216       LRR 7.
FT   REPEAT      217    240       LRR 8.
FT   REPEAT      242    264       LRR 9.
FT   REPEAT      266    288       LRR 10.
FT   REPEAT      289    312       LRR 11.
FT   REPEAT      313    336       LRR 12.
FT   REPEAT      337    360       LRR 13.
FT   REPEAT      361    384       LRR 14.
FT   REPEAT      386    408       LRR 15.
FT   REPEAT      409    432       LRR 16.
FT   REPEAT      433    456       LRR 17.
FT   REPEAT      458    478       LRR 18.
FT   REPEAT      479    504       LRR 19.
FT   REPEAT      506    529       LRR 20.
FT   REPEAT      543    566       LRR 21.
FT   CARBOHYD     64     64       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     85     85       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     96     96       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    368    368       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    515    515       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    586    586       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   603 AA;  66811 MW;  DCD7637D94A5037C CRC64;
```

```
  Query Match             9.3%;  Score 312;  DB 1;  Length 603;
  Best Local Similarity   30.0%;  Pred. No. 6.8e-14;
  Matches   95;  Conservative   46;  Mismatches  136;  Indels   40;  Gaps    5;

Qy      25 PLLVVLLALQLLV-----------VAGLVRAQTCPSVCSCSN----QFSKVICVRKNLRE 69
           | ||||||  : :             :       ||  |:||:        | |  |||
Db       8 PALVVLLAFWVALGPCHLQGTDPGASADAEGPQCPVACTCSHDDYTDELSVFCSSKNLTH 67

Qy      70 VPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTL 129
           :|| |  :|| | |  | : |   :|::|  |: | |  : :|::|  |  || ||  |
Db      68 LPDDIPVSTRALWLDGNNLSSIPSAAFQNLSSLDFLNLQGSWLRSLEPQALLGLQNLYYL 127

Qy     130 ELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLG--------- 180
           |  |||  :  | | :   |  | | :| :  :    |  :  |  |:||
Db     128 HLERNRLRNLAVGLFTHTPSLASLSLSSNLLGRLEEGLFQGLSHLWDLNLGWNSLVVLPD 187

Qy     181 ----------EL----KRLSYISEGAFEGLSNLRYLNLAMCNLREIPN--LTPLIKLDEL 224
                     ||     :|:|:    | ||  || |:|:   || :       | :| :|
Db     188 TVFQGLGNLHELVLAGNKLTYLQPALFCGLGELRELDLSRNALRSVKANVFVHLPRLQKL 247

Qy     225 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 284
           |  | ::|: ||:| |:  |: | :  :::  :  : |  |  |  : |||| :  |
Db     248 YLDRNLITAVAPGAFLGMKALRWLDLSHNRVAGLMEDTFPGLLGLHVLRLAHNAIASLRP 307

Qy     285 DLFTPLHHLERIHLHHN 301
             |  || || : | ||
Db     308 RTFKDLHFLEELQLGHN 324
```

```
RESULT  10
ALS_MOUSE
ID   ALS_MOUSE      STANDARD;      PRT;   603 AA.
AC   P70389;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Insulin-like growth factor binding protein complex acid labile chain
DE   precursor (ALS).
GN   IGFALS OR ALS OR ALBS.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SV;
RX   MEDLINE=96413591; PubMed=8816745;
RA   Boisclair Y.R., Seto D., Hsieh S., Hurst K.R., Ooi G.T.;
RT   "Organization and chromosomal localization of the gene encoding the
RT   mouse acid labile subunit of the insulin-like growth factor binding
RT   complex.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:10028-10033(1996).
```

```
CC   -!- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN REGULATING THE ACCESS OF
CC       CIRCULATING IGFS TO THE TISSUES.
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 kDa WITH
CC       IGF-I OR IGF-II AND IGFBP-3 (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U66900; AAB17270.1; -.
DR   MGD; MGI:107973; Igfals.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 19.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00370; LRR; 4.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 9.
KW   Glycoprotein; Leucine-rich repeat; Repeat; Signal.
FT   SIGNAL        1     23       BY SIMILARITY.
FT   CHAIN        24    603       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                 PROTEIN COMPLEX ACID LABILE CHAIN.
FT   REPEAT       52     73       LRR 1.
FT   REPEAT       74     96       LRR 2.
FT   REPEAT       98    120       LRR 3.
FT   REPEAT      121    144       LRR 4.
FT   REPEAT      146    168       LRR 5.
FT   REPEAT      169    192       LRR 6.
FT   REPEAT      194    216       LRR 7.
FT   REPEAT      217    240       LRR 8.
FT   REPEAT      242    264       LRR 9.
FT   REPEAT      265    288       LRR 10.
FT   REPEAT      289    312       LRR 11.
FT   REPEAT      313    336       LRR 12.
FT   REPEAT      338    360       LRR 13.
FT   REPEAT      361    384       LRR 14.
FT   REPEAT      386    408       LRR 15.
FT   REPEAT      409    432       LRR 16.
FT   REPEAT      433    456       LRR 17.
FT   REPEAT      458    480       LRR 18.
FT   REPEAT      482    504       LRR 19.
FT   REPEAT      505    529       LRR 20.
FT   REPEAT      543    566       LRR 21.
FT   CARBOHYD     64     64       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     85     85       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     96     96       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    368    368       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    515    515       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    586    586       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   603 AA;  66959 MW;  4A6223AEC7026BCC CRC64;


  Query Match             9.3%;  Score 311;  DB 1;  Length 603;
  Best Local Similarity   30.5%;  Pred. No. 8e-14;
  Matches  100;  Conservative   44;  Mismatches  122;  Indels   62;  Gaps    8;

Qy      25 PLLVVLLALQLLV-----------VAGLVRAQTCPSVCSCS-----NQFSKVICVRKNLR 68
           | ||||||  : :           :       ||  |:||     :: | | |  :||
Db       8 PALVVLLAFWVALGPCYLQGTDPGASADAEGPQCPVTCTCSYDDYTDELS-VFCSSRNLT 66

Qy      69 EVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNT 128
           ::||||  :|| | |  | :  |   :|:::|  |: | |  : :|::|  |  || ||
Db      67 QLPDGIPVSTRALWLDGNNLSSIPSAAFQNLSSLDFLNLQGSWLRSLEPQALLGLQNLYH 126

Qy     129 LELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESISPSYAFNRIPSLRRLDLGELKRLSYI 188
           | |  | | ::  | | :   |  | | || :  :    |  :  |  |:||    |  :
Db     127 LHLERNLLRSLAAGLFRHTPSLASLSLGNNLLGRLEEGLFRGLSHLWDLNLG-WNSLVVL 185

Qy     189 SEGAFEGLSNLRYLNLA-----------MCNLREIPNLTPLIKLDELDLSGNHLSAIRPG 237
             :  |:|| ||  | ||           :| | |         | |||| | | :::
Db     186 PDTVFQGLGNLHELVLAGNKLTYLQPALLCGLGE---------LRELDLSRNALRSVKAN 236

Qy     238 SFQGLMHLQKLWMIQSQIQVIERNAF-------------DNLQSLVE-----------IN 273
           |  |  ||||:: :: |  :   ||              : :  |:|           :
Db     237 VFIHLPRLQKLYLDRNLITAVAPRAFLGMKALRWLDLSHNRVAGLLEDTFPGLLGLHVLR 296

Qy     274 LAHNNLTLLPHDLFTPLHHLERIHLHHN 301
           |||| :| |    |  || || : | ||
Db     297 LAHNAITSLRPRTFKDLHFLEELQLGHN 324


RESULT  11
PGS2_HUMAN
ID   PGS2_HUMAN     STANDARD;      PRT;   359 AA.
AC   P07585; Q9Y5N9; Q9Y5N8; Q9P0Z0; Q9P0Z1;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-APR-1988 (Rel. 07, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Bone proteoglycan II precursor (PG-S2) (Decorin) (PG40).
GN   DCN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87017013; PubMed=3484330;
RA   Krusius T., Ruoslahti E.;
RT   "Primary structure of an extracellular matrix proteoglycan core
RT   protein deduced from cloned cDNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:7683-7687(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=93162643; PubMed=8432527;
RA   Vetter U., Vogel W., Just W., Young M.F., Fisher L.W.;
RT   "Human decorin gene: intron-exon junctions and chromosomal
RT   localization.";
RL   Genomics 15:161-168(1993).
RN   [3]
RP   SEQUENCE OF 1-70 FROM N.A.
RX   MEDLINE=93162642; PubMed=8432526;
RA   Danielson K.G., Fazzio A., Cohen I.R., Cannizzaro L., Iozzo R.V.;
RT   "The human decorin gene: intron-exon organization, discovery of two
RT   alternatively spliced exons in the 5' untranslated region, and
RT   mapping of the gene to chromosome 12q23.";
RL   Genomics 15:146-160(1993).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORMS A; B; C; D AND E).
RA   Cs-Szabo G., Glant T.T.;
RT   "Alternative splicing of human decorin.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 31-50.
RX   MEDLINE=90073579; PubMed=2590169;
RA   Roughley P.J., White R.J.;
RT   "Dermatan sulphate proteoglycans of human articular cartilage. The
RT   properties of dermatan sulphate proteoglycans I and II.";
RL   Biochem. J. 262:823-827(1989).
RN   [6]
RP   SEQUENCE OF 31-49.
RX   MEDLINE=87250639; PubMed=3597437;
RA   Fisher L.W., Hawkins G.R., Tuross N., Termine J.D.;
```

```
RT   "Purification and partial characterization of small proteoglycans I
RT   and II, bone sialoproteins I and II, and osteonectin from the mineral
RT   compartment of developing human bone.";
RL   J. Biol. Chem. 262:9702-9708(1987).
CC   -!- FUNCTION: BINDS TO TYPE I AND TYPE II COLLAGEN AND AFFECTS THE
CC       RATE OF FIBRILS FORMATION. ALSO BINDS TO FIBRONECTIN AND TGF-
CC       BETA.
CC   -!- PTM: THE GLYCOSAMINOGLYCAN CHAIN ATTACHED TO DECORIN CAN BE EITHER
CC       CHONDROITIN SULFATE OR DERMATAN SULFATE DEPENDING UPON THE
CC       TISSUE OF ORIGIN.
CC   -!- ALTERNATIVE PRODUCTS: 5 isoforms; A (shown here), B, C, D and E;
CC       are produced by alternative splicing.
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 10 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M14219; AAB00774.1; -.
DR   EMBL; L01131; AAA52301.1; ALT_SEQ.
DR   EMBL; L01125; AAA52301.1; JOINED.
DR   EMBL; L01126; AAA52301.1; JOINED.
DR   EMBL; L01127; AAA52301.1; JOINED.
DR   EMBL; L01129; AAA52301.1; JOINED.
DR   EMBL; L01130; AAA52301.1; JOINED.
DR   EMBL; M98262; AAB60901.1; -.
DR   EMBL; AF138300; AAD44713.1; -.
DR   EMBL; AF138301; AAF61437.1; -.
DR   EMBL; AF138302; AAD44714.1; -.
DR   EMBL; AF138303; AAF61438.1; -.
DR   EMBL; AF138304; AAD44715.1; -.
DR   PIR; A26476; NBHUC8.
DR   PIR; S05640; S05640.
DR   PIR; B28457; B28457.
DR   PIR; A45016; A45016.
DR   MIM; 125255; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 9.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00370; LRR; 3.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 1.
KW   Glycoprotein; Connective tissue; Extracellular matrix; Proteoglycan;
KW   Repeat; Leucine-rich repeat; Signal; Alternative splicing;
KW   Polymorphism.
FT   SIGNAL       1    16       POTENTIAL.
FT   PROPEP      17    30
FT   CHAIN       31   359       BONE PROTEOGLYCAN II.
FT   REPEAT      77    98       LRR 1.
FT   REPEAT      99   122       LRR 2.
FT   REPEAT     123   145       LRR 3.
FT   REPEAT     146   167       LRR 4.
FT   REPEAT     168   193       LRR 5.
FT   REPEAT     194   217       LRR 6.
FT   REPEAT     218   238       LRR 7.
FT   REPEAT     239   262       LRR 8.
FT   REPEAT     263   285       LRR 9.
FT   REPEAT     286   308       LRR 10.
FT   CARBOHYD    34    34       O-LINKED (GLYCOSAMINOGLYCAN).
FT   CARBOHYD   211   211       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   262   262       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   303   303       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    54    67       BY SIMILARITY.
FT   DISULFID   313   346       BY SIMILARITY.
FT   VARSPLIC    71   179       MISSING (IN ISOFORM B).
FT   VARSPLIC    73   219       MISSING (IN ISOFORM C).
FT   VARSPLIC   109   295       MISSING (IN ISOFORM D).
FT   VARSPLIC    72    75       LDKV -> CLPS (IN ISOFORM E).
FT   VARSPLIC    76   359       MISSING (IN ISOFORM E).
FT   VARIANT    273   273       E -> Q (IN DBSNP:1803344).
FT                              /FTId=VAR_011975.
FT   CONFLICT    37    37       G -> A (IN REF. 6).
FT   CONFLICT    45    45       D -> P (IN REF. 6).
SQ   SEQUENCE   359 AA;  39746 MW;  FF511E871A1A52DD CRC64;
```

```
  Query Match             9.0%;  Score 302;  DB 1;  Length 359;
  Best Local Similarity   33.1%;  Pred. No. 1.7e-13;
  Matches   85;  Conservative   41;  Mismatches  121;  Indels   10;  Gaps    6;

Qy      47 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQ 106
           ||  | |      | |    | :||  :  :| ||:|  |:|  ||   ||:|::|  |
Db      54 CPFRCQC--HLRVVQCSDLGLDKVPKDLPPDTTLLDLQNNKITEIKDGDFKNLKNLHALI 111

Qy     107 LSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSY 166
           |  | |  :  |||  |  |  | |  |:|  :|          |:||    | |  :
Db     112 LVNNKISKVSPGAFTPLVKLERLYLSKNQLKELPE---KMPKTLQELRAHENEITKVRKV 168

Qy     167 AFNRIPSLRRLDLG--ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 224
           || :  :  ::||   ||  | |  |||:|:  | |: :|  |:  ||   |   | ||
Db     169 TFNGLNQMIVIELGTNPLKS-SGIENGAFQGMKKLSYIRIADTNITSIPQGLP-PSLTEL 226

Qy     225 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 284
           | || :| :   | :|| :| || :  : |  ::  :  |   | |::| :| || :|
Db     227 HLDGNKISRVDAASLKGLNNLAKLGLSFNSISAVDNGSLANTPHLRELHLDNNKLTRVPG 286

Qy     285 DLFTPLHHLERIHLHHN 301
           |     ::: ::||:|
Db     287 GL-AEHKYIQVVYLHNN 302
```

```
RESULT  12
PGS2_HORSE
ID   PGS2_HORSE     STANDARD;      PRT;   360 AA.
AC   O46542;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Bone proteoglycan II precursor (PG-S2) (Decorin) (Dermatan sulfate
DE   proteoglycan II) (DS-PGII).
GN   DCN.
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Richardson D.W., Dodge G.R.;
RT   "Effects of interleukin-1 beta and tumor necrosis factor-alpha on the
RT   expression of matrix related genes in cultured equine articular
RT   chondrocytes.";
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS TO TYPE I AND TYPE II COLLAGEN AND AFFECTS THE
CC       RATE OF FIBRILS FORMATION. ALSO BINDS TO FIBRONECTIN AND TGF-
CC       BETA (BY SIMILARITY).
CC   -!- PTM: THE GLYCOSAMINOGLYCAN CHAIN ATTACHED TO DECORIN CAN BE EITHER
CC       CHONDROITIN SULFATE OR DERMATAN SULFATE DEPENDING UPON THE
CC       TISSUE OF ORIGIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 10 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF038127; AAB92652.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   Pfam; PF00560; LRR; 9.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00370; LRR; 3.
DR   SMART; SM00013; LRRNT; 1.
KW   Glycoprotein; Connective tissue; Extracellular matrix; Proteoglycan;
KW   Repeat; Leucine-rich repeat; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   PROPEP       17     30       BY SIMILARITY.
FT   CHAIN        31    360       BONE PROTEOGLYCAN II.
FT   REPEAT       78     99       LRR 1.
FT   REPEAT      100    123       LRR 2.
FT   REPEAT      124    146       LRR 3.
FT   REPEAT      147    168       LRR 4.
FT   REPEAT      169    194       LRR 5.
FT   REPEAT      195    218       LRR 6.
FT   REPEAT      219    239       LRR 7.
FT   REPEAT      240    263       LRR 8.
FT   REPEAT      264    286       LRR 9.
FT   REPEAT      287    309       LRR 10.
FT   CARBOHYD     34     34       O-LINKED (GLYCOSAMINOGLYCAN) (BY
FT                                SIMILARITY).
FT   CARBOHYD    190    190       O-LINKED (GLYCOSAMINOGLYCAN) (BY
FT                                SIMILARITY).
FT   CARBOHYD    212    212       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    263    263       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    304    304       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID     55     68       BY SIMILARITY.
FT   DISULFID    314    347       BY SIMILARITY.
SQ   SEQUENCE   360 AA;  39939 MW;  2DAE97CDE16F7C45 CRC64;


  Query Match             8.9%;  Score 299;  DB 1;  Length 360;
  Best Local Similarity   29.4%;  Pred. No. 2.7e-13;
  Matches  107;  Conservative   47;  Mismatches  150;  Indels   60;  Gaps   13;

Qy      15 GPRFNRALFDPLLV----------VLLALQLLVVAGLVRAQTCPSVCSCSNQFSKVI 61
           ||   | ||| :|              :| |: |          ||  | |      |
Db      17 GPFQQRGLFDFMLEDEASGIGPEDRIHEVLDLEPL-------GPVCPFRCQC--HLRVVQ 67

Qy      62 CVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFN 121
           |    | :||  :  :| ||:|  |:|  ||   ||:|::|  | |  | |  |  |||
Db      68 CSDLGLDKVPKDLPPDTTLLDLQNNKITEIKDGDFKNLKNLHALILVNNKISKISPGAFT 127

Qy     122 GLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLG- 180
           |  | | | | | :|         |:|| :  | | :    || :  :  ::|| |
Db     128 PLVKLERLYLSKNHLKELPE---KMPKTLQELRVHENEITKVRKAVFNGLNQMIVVELGT 184

Qy     181 -ELKRLSYISEGAFEGLSNLRYLNLAMCNLREI-PNLTPLIKLDELDLSGNHLSAIRPGS 238
            ||  | |  |||:|:  | |: :|  |:  | | | |   | || | || :: :   |
Db     185 NPLKS-SGIENGAFQGMKKLSYIRIADTNITTIPPGLPP--SLTELHLDGNKITKVDAAS 241

Qy     239 FQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 298
           :|| :| || :  : |  ::  :  |   | |::| :| |  :|  |     ::: ::|
Db     242 LRGLNNLAKLGLSFNSISAVDNGSLANTPHLRELHLDNNKLIKVPGGL-ADHKYIQVVYL 300

Qy     299 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIG----------ELDQN 347
           |:|  | :               ||  |    ||    |  | |           |:  : :
Db     301 HNN--NISA-----------VGSNDFCPPGYNTK---KASYSGVSLFSNPVQYWEIQPS 343

Qy     348 YFTC 351
           | |
Db     344 TFRC 347


RESULT  13
GPBA_HUMAN

ID   GPBA_HUMAN     STANDARD;      PRT;   626 AA.
AC   P07359;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-APR-1988 (Rel. 07, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Platelet glycoprotein IB alpha chain precursor (GP-IB alpha) (GPIBA)
DE   (CD42B-alpha) (CD42B) [Contains: Glycocalicin].
GN   GP1BA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=87289655; PubMed=3303030;
RA   Lopez J.A., Chung D.W., Fujikawa K., Hagen F.S., Papayannopoulou T.,
RA   Roth G.J.;
RT   "Cloning of the alpha chain of human platelet glycoprotein Ib: a
RT   transmembrane protein with homology to leucine-rich alpha 2-
RT   glycoprotein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:5615-5619(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89025874; PubMed=2845978;
RA   Wenger R.H., Kieffer N., Wicki A.N., Clemetson K.J.;
RT   "Structure of the human blood platelet membrane glycoprotein Ib alpha
RT   gene.";
RL   Biochem. Biophys. Res. Commun. 156:389-395(1988).
RN   [3]
RP   SEQUENCE OF 17-315.
RX   MEDLINE=87289654; PubMed=3497398;
RA   Titani K., Takio K., Handa M., Ruggeri Z.M.;
RT   "Amino acid sequence of the von Willebrand factor-binding domain of
RT   platelet membrane glycoprotein Ib.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:5610-5614(1987).
RN   [4]
RP   DISULFIDE BONDS.
RX   MEDLINE=91301149; PubMed=2070794;
RA   Hess D., Schaller J., Rickli E.E., Clemetson K.J.;
RT   "Identification of the disulphide bonds in human platelet
RT   glycocalicin.";
RL   Eur. J. Biochem. 199:389-393(1991).
RN   [5]
RP   VARIANT SIBA.
RX   MEDLINE=92265982; PubMed=1586750;
RA   Murata M., Furihata K., Ishida F., Russell S.R., Ware J.,
RA   Ruggeri Z.M.;
RT   "Genetic and structural characterization of an amino acid dimorphism
RT   in glycoprotein Ib alpha involved in platelet transfusion
RT   refractoriness.";
RL   Blood 79:3086-3090(1992).
RN   [6]
RP   VARIANT BSS PHE-73.
RX   MEDLINE=92110577; PubMed=1730088;
RA   Miller J.L., Lyle V.A., Cunningham D.;
RT   "Mutation of leucine-57 to phenylalanine in a platelet glycoprotein
RT   Ib alpha leucine tandem repeat occurring in patients with an
RT   autosomal dominant variant of Bernard-Soulier disease.";
RL   Blood 79:439-446(1992).
RN   [7]
RP   VARIANT BSS VAL-172.
RX   MEDLINE=93388851; PubMed=7690774;
RA   Ware J., Russell S.R., Marchese P., Murata M., Mazzucato M.,
RA   de Marco L., Ruggeri Z.M.;
RT   "Point mutation in a leucine-rich repeat of platelet glycoprotein Ib
RT   alpha resulting in the Bernard-Soulier syndrome.";
RL   J. Clin. Invest. 92:1213-1220(1993).
RN   [8]
RP   VARIANT BSS SER-225.
RX   MEDLINE=95118882; PubMed=7819107;
RA   Simsek S., Noris P., Lozano M., Pico M., von Dem Borne A.E.G.K.,
RA   Ribera A., Gallardo D.;
RT   "Cys209 Ser mutation in the platelet membrane glycoprotein Ib alpha
```

```
RT   gene is associated with Bernard-Soulier syndrome.";
RL   Br. J. Haematol. 88:839-844(1994).
RN   [9]
RP   VARIANT PSEUDO-VWD VAL-249.
RX   MEDLINE=91271273; PubMed=2052556;
RA   Miller J.L., Cunningham D., Lyle V.A., Finch C.N.;
RT   "Mutation in the gene encoding the alpha chain of platelet
RT   glycoprotein Ib in platelet-type von Willebrand disease.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:4761-4765(1991).
RN   [10]
RP   VARIANT PSEUDO-VWD VAL-249.
RX   MEDLINE=93253059; PubMed=8486780;
RA   Murata M., Russell S.R., Ruggeri Z.M., Ware J.;
RT   "Expression of the phenotypic abnormality of platelet-type von
RT   Willebrand disease in a recombinant glycoprotein Ib alpha fragment.";
RL   J. Clin. Invest. 91:2133-2137(1993).
RN   [11]
RP   VARIANT PSEUDO-VWD VAL-255.
RX   MEDLINE=93214031; PubMed=8384898;
RA   Russell S.D., Roth G.J.;
RT   "Pseudo-von Willebrand disease: a mutation in the platelet
RT   glycoprotein Ib alpha gene associated with a hyperactive surface
RT   receptor.";
RL   Blood 81:1787-1791(1993).
RN   [12]
RP   VARIANT BSS LEU-195 DEL.
RX   MEDLINE=95178321; PubMed=7873390;
RA   de la Salle C., Baas M.-J., Lanza F., Schwartz A., Hanau D.,
RA   Chevalier J., Gachet C., Briquel M.-E., Cazenave J.-P.;
RT   "A three-base deletion removing a leucine residue in a leucine-rich
RT   repeat of platelet glycoprotein Ib alpha associated with a variant of
RT   Bernard-Soulier syndrome (Nancy I).";
RL   Br. J. Haematol. 89:386-396(1995).
CC   -!- FUNCTION: GP-IB, A SURFACE MEMBRANE PROTEIN OF PLATELETS,
CC       PARTICIPATES IN THE FORMATION OF PLATELET PLUGS BY BINDING TO VON
CC       WILLEBRAND FACTOR, WHICH IS ALREADY BOUND TO THE SUBENDOTHELIUM.
CC   -!- SUBUNIT: GP-IB ALPHA AND BETA ARE DISULFIDE LINKED. GP-IX IS
CC       COMPLEXED WITH THE GP-IB HETERODIMER VIA A NON COVALENT LINKAGE.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: GLYCOCALICIN, WHICH IS APPROXIMATELY COEXTENSIVE WITH THE
CC       EXTRACELLULAR PART OF THE MOLECULE, IS CLEAVED OFF BY CALPAIN
CC       DURING PLATELET LYSIS.
CC   -!- POLYMORPHISM: POSITION 161 IS ASSOCIATED WITH PLATELET-SPECIFIC
CC       ALLOANTIGEN SIBA. SIBA(-) HAS THR-161 AND SINA(+) HAS MET-161.
CC       SIBA IS INVOLVED IN NEONATAL ALLOIMMUNE THROMBOCYTOPENIA (NATP).
CC   -!- DISEASE: DEFECTS IN GP1BA ARE ONE OF THE CAUSES OF BERNARD-SOULIER
CC       SYNDROME (BSS). BSS PATIENTS HAVE UNUSUALLY LARGE PLATELETS AND
CC       HAVE A CLINICAL BLEEDING TENDENCY.
CC   -!- DISEASE: DEFECTS IN GP1BA ARE ONE OF THE CAUSES OF VON WILLEBRAND
CC       DISEASE (VWD) KNOWN AS PLATELET-TYPE VON WILLEBRAND DISEASE OR
CC       PSEUDO-VON WILLEBRAND DISEASE. THIS AUTOSOMAL DOMINANT BLEEDING
CC       DISORDER IS CAUSED BY AN INCREASED AFFINITY OF GP-IB FOR SOLUBLE
CC       VWF RESULTING IN IMPAIRED HEMOSTATIC FUNCTION DUE TO THE REMOVAL
CC       OF VWF FROM THE CIRCULATION.
CC   -!- MISCELLANEOUS: PLATELET ACTIVATION APPARENTLY INVOLVES DISRUPTION
CC       OF THE MACROMOLECULAR COMPLEX OF GP-IB WITH THE PLATELET
CC       GLYCOPROTEIN IX (GP-IX) AND DISSOCIATION OF GP-IB FROM THE
CC       ACTIN-BINDING PROTEIN.
CC   -!- MISCELLANEOUS: BINDING SITES FOR VON WILLEBRAND FACTOR AND
CC       THROMBIN (THE LATTER SITE WITH UNKNOWN FUNCTION) ARE IN THE
CC       AMINO-TERMINAL PART OF THE MOLECULE.
CC   -!- SIMILARITY: CONTAINS 6 LEUCINE-RICH REPEATS (LRR).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; J02940; AAA52595.1; -.
DR   EMBL; M22403; AAA52596.1; -.
DR   PIR; A27075; NBHUIA.
DR   GlycoSuiteDB; P07359; -.
DR   MIM; 231200; -.
DR   MIM; 177820; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 6.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00370; LRR; 3.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 2.
KW   Platelet; Transmembrane; Glycoprotein; Hemostasis; Blood coagulation;
KW   Repeat; Leucine-rich repeat; Signal; Cell adhesion; Disease mutation;
KW   Polymorphism; von Willebrand disease; Bernard Soulier syndrome.
FT   SIGNAL        1     16
FT   CHAIN        17    626       PLATELET GLYCOPROTEIN IB ALPHA CHAIN.
FT   CHAIN        17      ?       GLYCOCALICIN.
FT   DOMAIN       17    505       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    506    526       POTENTIAL.
FT   DOMAIN      527    626       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       70     92       LRR 1.
FT   REPEAT       93    117       LRR 2.
FT   REPEAT      119    138       LRR 3.
FT   REPEAT      139    162       LRR 4.
FT   REPEAT      164    186       LRR 5.
FT   REPEAT      188    210       LRR 6.
FT   REPEAT      379    386       THR/PRO-RICH.
FT   REPEAT      387    395       THR/PRO-RICH.
FT   REPEAT      400    408       THR/PRO-RICH.
FT   REPEAT      409    417       THR/PRO-RICH.
FT   REPEAT      422    430       THR/PRO-RICH.
FT   DISULFID     20     33
FT   DISULFID    225    264
FT   DISULFID    227    280
FT   CARBOHYD     37     37       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    175    175       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    308    308       O-LINKED.
FT   VARIANT      72     72       R -> H (IN DBSNP:6068).
FT                                /FTId=VAR_011909.
FT   VARIANT      73     73       L -> F (IN BSS).
FT                                /FTId=VAR_005256.
FT   VARIANT     161    161       T -> M (IN ALLOANTIGEN SIBA(+)).
FT                                /FTId=VAR_005257.
FT   VARIANT     172    172       A -> V (IN BSS).
FT                                /FTId=VAR_005258.
FT   VARIANT     195    195       MISSING (IN BSS).
FT                                /FTId=VAR_005259.
FT   VARIANT     225    225       C -> S (IN BSS).

  Query Match             8.8%;  Score 296.5;  DB 1;  Length 626;
  Best Local Similarity   24.2%;  Pred. No. 8e-13;
  Matches  150;  Conservative   81;  Mismatches  193;  Indels  197;  Gaps   27;

Qy      25 PLLVVLLALQLLVVAGLVRAQTCPS------VCSCSNQFS--KVICVRKNLREVPDGIST 76
           |||::|| |              ||      :|  |   |  :| | ::||  :|  :
Db       2 PLLLLLLLL--------------PSPLHPHPICEVSKVASHLEVNCDKRNLTALPPDLPK 47

Qy      77 NTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRL 136
           :| :|:| || :    : :       |  | | |  :  :::     |  | ||:|  |:|
Db      48 DTTILHLSENLLYTFSLATLMPYTRLTQLNLDRCELTKLQVD--GTLPVLGTLDLSHNQL 105

Qy     137 TTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGL 196
             ::|            |  :  |  ::   :|||: ||               ||  ||
Db     106 QSLP------------LLGQTLPALTVLDVSFNRLTSL-------------PLGALRGL 139

Qy     197 SNLRYLNLAMCNLREIPN--LTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQ 254
             |: | |    |: :|   |||  ||::| |: |:|: :  |   ||
```

```
Db    140 GELQELYLKGNELKTLPPGLLTPTPKLEKLSLANNNLTELPAGLLNGL------------ 187

Qy    255 IQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSW 314
                :|| :|:   |  |:|  :|   |   | |    || ||| |||:||:
Db    188 ---------ENLDTLL---LQENSLYTIPKGFFGS-HLLPFAFLHGNPWLCNCEILYFRR 234

Qy    315 WIKDMA---------------PSNTACCARCNTPPNLKGRYIG--------ELDQNYFTC 351
          |::| |               || |     |:      :| |        | | : :
Db    235 WLQDNAENVYVWKQGVDVKAMTSNVASVQCDNSDKFPVYKYPGKGCPTLGDEGDTDLYDY 294

Qy    352 YAPVIVEPPADLNVTEGMAAELKCRASTSLTS-------------VSWIT-------PNG 391
          |      |  |   |||     | ||: ::                 || |        |:
Db    295 Y------PEED---TEGD----KVRATRTVVKFPTKAHTTPWGLFYSWSTASLDSQMPSS 341

Qy    392 TVMTHGAYKVRIAVLSDGTLNFT----NVTVQDTGMYTCMVSNSVGNTTAS--------- 438
            |  : | :       | |||    ::|   |   |   : |   ||:
Db    342 LHPTQESTKEQTTFPPRWTPNFTLHMESITFSKTPKSTTEPTPS--PTTSEPVPEPAPNM 399

Qy    439 ATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTPVVDWETT-----NVTTSL 493
           ||  | : |||            ||: : |  :     |||:    |:     : |: :
Db    400 TTLEPTPSPTTP------------EPTSEPA-PSPTTPEPTPIPTIATSPTILVSATSLI 446

Qy    494 TPQSTRSTEKTFTIPVTDINS---GIPGIDEVMKTTKIIIG----------------CFV 534
          ||   :||  | | ||: : |    || :|:  |   :: |                | :
Db    447 TP---KSTFLTTTKPVSLLESTKKTIPELDQPPKLRGVLQGHLESSRNDPFLHPDFCCLL 503

Qy    535 AI---------TLMAAVMLVI 546
            :          | |:|:|::
Db    504 PLGFYVLGLFWLLFASVVLIL 524


RESULT  14
PGS2_CANFA
ID   PGS2_CANFA     STANDARD;      PRT;   360 AA.
AC   Q29393;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Bone proteoglycan II precursor (PG-S2) (Decorin).
GN   DCN OR DCN1C.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Glant T.T.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 244-259 FROM N.A.
RA   Venta P.J., Brouillette J.A., Yuzbasiyan-Gurkan V., Brewer G.J.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS TO TYPE I AND TYPE II COLLAGEN AND AFFECTS THE
CC       RATE OF FIBRILS FORMATION. ALSO BINDS TO FIBRONECTIN AND TGF-
CC       BETA (BY SIMILARITY).
CC   -!- PTM: THE GLYCOSAMINOGLYCAN CHAIN ATTACHED TO DECORIN CAN BE EITHER
CC       CHONDROITIN SULFATE OR DERMATAN SULFATE DEPENDING UPON THE
CC       TISSUE OF ORIGIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 10 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U83141; AAB51245.1; -.
DR   EMBL; L77684; AAA98062.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 9.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00370; LRR; 2.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 1.
KW   Glycoprotein; Connective tissue; Extracellular matrix; Proteoglycan;
KW   Repeat; Leucine-rich repeat; Signal.
FT   SIGNAL       1    16       POTENTIAL.
FT   PROPEP      17    30       BY SIMILARITY.
FT   CHAIN       31   360       BONE PROTEOGLYCAN II.
FT   REPEAT      78    99       LRR 1.
FT   REPEAT     100   123       LRR 2.
FT   REPEAT     124   146       LRR 3.
FT   REPEAT     147   168       LRR 4.
FT   REPEAT     169   194       LRR 5.
FT   REPEAT     195   218       LRR 6.
FT   REPEAT     219   239       LRR 7.
FT   REPEAT     240   263       LRR 8.
FT   REPEAT     264   286       LRR 9.
FT   REPEAT     287   309       LRR 10.
FT   CARBOHYD    34    34       O-LINKED (GLYCOSAMINOGLYCAN) (BY
FT                               SIMILARITY).
FT   CARBOHYD   212   212       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   263   263       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   304   304       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    55    68       BY SIMILARITY.
FT   DISULFID   314   347       BY SIMILARITY.
SQ   SEQUENCE   360 AA;  39980 MW;  99BEE11A9C812906 CRC64;


  Query Match             8.8%;  Score 295;  DB 1;  Length 360;
  Best Local Similarity   29.2%;  Pred. No. 5e-13;
  Matches  106;  Conservative   45;  Mismatches  154;  Indels   58;  Gaps   12;

Qy     15 GPRFNRALFDPLL-------------VVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKVI 61
          ||   | ||| :|               :  |:||          ||  | |      |
Db     17 GPFQQRGLFDFMLEDEASGIGPEDRAPDMPDLELL-------GPVCPFRCQC--HLRVVQ 67

Qy     62 CVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFN 121
          |    | :||  :  :| ||:|  |:|  ||   ||:|::|  |  |  |  |  |||
Db     68 CSDLGLDKVPKDLPPDTTLLDLQNNKITEIKDGDFKNLKNLHTLILVNNKISKISPGAFT 127

Qy    122 GLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLG- 180
            |  |  | |  | |  :|         |:||    | |  :    || :  :  ::||
Db    128 PLLKLERLYLSKNHLKELPE---KMPKTLQELRAHENEITKVRKAVFNGLNQMIVVELGT 184

Qy    181 -ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSF 239
           ||  | |  |||:|:  | |: :|  |:  ||   |   | || | || :: :   |
Db    185 NPLKS-SGIENGAFQGMKKLSYIRIADTNITTIPQGLP-PSLTELHLEGNKITKVDASSL 242

Qy    240 QGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLH 299
          :|| :| || :  : |  ::      |  | |::| :| |  :|  |     ::: ::||
Db    243 KGLNNLAKLGLSFNSISAVDNGTLANTPHLRELHLDNNKLIRVPGGL-AEHKYIQVVYLH 301

Qy    300 HNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIG----------ELDQNY 348
          :|  | :              || |    ||    |  | |           |:  :
Db    302 NN--NISA------------VGSNDFCPPGYNTK---KASYSGVSLFSNPVQYWEIQPST 344

Qy    349 FTC 351
          | |
Db    345 FRC 347


RESULT  15
PGS2_PIG
ID   PGS2_PIG       STANDARD;      PRT;   360 AA.
AC   Q9XSD9; Q9XSH4;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
```

```
DE   Bone proteoglycan II precursor (PG-S2) (Decorin).
GN   DCN.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A. (LONG FORM).
RC   STRAIN=YORKSHIRE;
RA   Stephenson S., Schnoke M., Vesely I.;
RT   "Cloning of the porcine decorin gene.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A. (SHORT FORM).
RC   STRAIN=YORKSHIRE; TISSUE=Aorta;
RA   Stephenson S., Schnoke M., Vesely I.;
RT   "Alternatively spliced version of the porcine decorin gene.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS TO TYPE I AND TYPE II COLLAGEN AND AFFECTS THE
CC       RATE OF FIBRILS FORMATION. ALSO BINDS TO FIBRONECTIN AND TGF-
CC       BETA (BY SIMILARITY).
CC   -!- PTM: THE GLYCOSAMINOGLYCAN CHAIN ATTACHED TO DECORIN CAN BE EITHER
CC       CHONDROITIN SULFATE OR DERMATAN SULFATE DEPENDING UPON THE
CC       TISSUE OF ORIGIN.
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 10 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF125537; AAD23578.1; -.
DR   EMBL; AF140270; AAD33862.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 9.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00370; LRR; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 2.
KW   Glycoprotein; Connective tissue; Extracellular matrix; Proteoglycan;
KW   Repeat; Leucine-rich repeat; Signal; Alternative splicing.
FT   SIGNAL        1     16       POTENTIAL.
FT   PROPEP       17     30       BY SIMILARITY.
FT   CHAIN        31    360       BONE PROTEOGLYCAN II.
FT   REPEAT       78     99       LRR 1.
FT   REPEAT      100    123       LRR 2.
FT   REPEAT      124    146       LRR 3.
FT   REPEAT      147    168       LRR 4.
FT   REPEAT      169    194       LRR 5.
FT   REPEAT      195    218       LRR 6.
FT   REPEAT      219    239       LRR 7.
FT   REPEAT      240    263       LRR 8.
FT   REPEAT      264    286       LRR 9.
FT   REPEAT      287    309       LRR 10.
FT   CARBOHYD     34     34       O-LINKED (GLYCOSAMINOGLYCAN) (BY
FT                                SIMILARITY).
FT   CARBOHYD    212    212       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    263    263       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    304    304       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID     55     68       BY SIMILARITY.
FT   DISULFID    314    347       POTENTIAL.
FT   VARSPLIC    281    318       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   360 AA;  39899 MW;  8573DE8DDEBA7509 CRC64;
```

```
  Query Match             8.8%;  Score 295;  DB 1;  Length 360;
  Best Local Similarity   30.5%;  Pred. No. 5e-13;
  Matches   97;  Conservative   44;  Mismatches  139;  Indels   38;  Gaps   10;

Qy      47 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQ 106
           ||  | |       | |    | :||  :  :| ||:|  |:|  ||   ||:|::|  |
Db      55 CPFRCQC--HLRVVQCSDLGLDKVPKDLPPDTALLDLQNNKITEIKDGDFKNLKNLHTLI 112

Qy     107 LSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSY 166
           |  | |  |  |||  |  |  | |  |:|  :|         |:|| :  | |  :
Db     113 LINNKISKISPGAFAPLVKLERLYLSKNQLKELPE---KMPKTLQELRVHENEITKVRKA 169

Qy     167 AFNRIPSLRRLDLG--ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 224
           || :  :  ::||   ||  | |  |||:|:  | |: :|  |:  ||   |   | ||
Db     170 VFNGLNQMIVVELGTNPLKS-SGIENGAFQGMKKLSYIRIADTNITTIPQGLP-PSLTEL 227

Qy     225 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 284
           | || :| :   | :|| :| || :  : |  ::  :  |   | |::| :| |  :|
Db     228 HLDGNKISKVDAASLKGLNNLAKLGLGFNSISTVDNGSLANTPHLRELHLNNNKLNKVPG 287

Qy     285 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIG-- 342
           |     ::: ::||:|  | :                ||  |    ||    |  | |
Db     288 GL-AEHKYIQVVYLHNN--NISA-----------VGSNDFCPPGYNTK---KASYSGVS 329

Qy     343 ---------ELDQNYFTC 351
                   |:  : | |
Db     330 LFSNPVQYWEIQPSTFRC 347


Search completed: August 26, 2002, 15:45:35
Job time: 247 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 2002, 15:41:03 ; Search time 34.22 Seconds
(without alignments)
3235.440 Million cell updates/sec

Title: US-09-905-056-292
Perfect score: 3362
Sequence: 1 MLNKMTLHPQQIMIGPRFNR.....VHEPLLIRMNSKDNVQETQI 640

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3362	100.0	640	4	Q9HCJ2	Q9hcj2 homo sapien
2	1865	55.5	653	4	Q9HBW1	Q9hbw1 homo sapien
3	1840	54.7	649	4	Q96A85	Q96a85 homo sapien
4	1376.5	40.9	422	4	Q9NT99	Q9nt99 homo sapien
5	1153	34.3	441	11	Q99PH1	Q99ph1 mus musculu
6	583.5	17.4	614	4	Q96FE5	Q96fe5 homo sapien
7	578.5	17.2	614	6	Q9N008	Q9n008 macaca fasc
8	577.5	17.2	614	11	Q9D1T0	Q9d1t0 mus musculu
9	575	17.1	606	4	Q9BZ20	Q9bz20 homo sapien
10	480.5	14.3	719	4	Q96NI6	Q96ni6 homo sapien
11	466	13.9	1021	5	Q9V430	Q9v430 drosophila
12	442.5	13.2	832	4	Q9ULH4	Q9ulh4 homo sapien
13	437.5	13.0	789	6	Q9BE71	Q9be71 macaca fasc
14	432	12.8	705	4	O43377	O43377 homo sapien
15	432	12.8	708	4	Q9H3W5	Q9h3w5 homo sapien
16	431.5	12.8	492	11	Q99KT6	Q99kt6 mus musculu
17	430	12.8	730	4	Q9P231	Q9p231 homo sapien
18	426.5	12.7	718	13	O73675	O73675 xenopus lae
19	426.5	12.7	788	11	Q9CYK3	Q9cyk3 mus musculu
20	423.5	12.6	716	11	Q61809	Q61809 mus musculu
21	420	12.5	1093	4	Q96JA1	Q96ja1 homo sapien
22	420	12.5	1094	4	Q9BYB8	Q9byb8 homo sapien
23	419	12.5	707	11	P97860	P97860 mus musculu
24	418	12.4	359	4	Q96RJ5	Q96rj5 homo sapien
25	417.5	12.4	1091	11	P70193	P70193 mus musculu
26	414	12.3	359	4	O15335	O15335 homo sapien
27	406	12.1	707	11	Q9ESY6	Q9esy6 rattus norv
28	401	11.9	674	4	Q9NZU1	Q9nzu1 homo sapien
29	401	11.9	1531	11	Q9WVB5	Q9wvb5 mus musculu
30	399.5	11.9	1065	4	O94898	O94898 homo sapien
31	399	11.9	1531	11	O88279	O88279 rattus norv
32	397.5	11.8	1521	4	O95710	O95710 homo sapien
33	397.5	11.8	1529	4	O94813	O94813 homo sapien
34	397	11.8	1496	4	Q92626	Q92626 homo sapien
35	395.5	11.8	358	11	O55226	O55226 mus musculu
36	395.5	11.8	1525	4	Q9Y5Q7	Q9y5q7 homo sapien
37	394.5	11.7	1521	11	Q9R1B9	Q9r1b9 mus musculu
38	394.5	11.7	1530	11	Q9WUG5	Q9wug5 rattus norv
39	393	11.7	358	11	O70210	O70210 rattus norv
40	389.5	11.6	1515	13	Q9DE37	Q9de37 brachydanio
41	387.5	11.5	1530	13	Q90WZ3	Q90wz3 xenopus lae
42	386.5	11.5	796	11	Q9WVC1	Q9wvc1 rattus norv
43	385	11.5	649	4	Q9NZU0	Q9nzu0 homo sapien
44	385	11.5	662	4	Q9P259	Q9p259 homo sapien
45	385	11.5	1534	4	O75093	O75093 homo sapien

ALIGNMENTS

RESULT 1
Q9HCJ2
ID Q9HCJ2 PRELIMINARY; PRT; 640 AA.
AC Q9HCJ2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA1580 PROTEIN (FRAGMENT).
GN KIAA1580.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20450683; PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirosawa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVIII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro.";
RL DNA Res. 7:273-281(2000).
DR EMBL; AB046800; BAB13406.1; -.
DR HSSP; P22888; 1LUT.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR003591; LRR_typ.
DR Pfam; PF00047; ig; 1.
DR Pfam; PF00560; LRR; 9.
DR Pfam; PF01463; LRRCT; 1.
DR Pfam; PF01462; LRRNT; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00409; IG; 1.
DR SMART; SM00408; IGc2; 1.

```
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00370; LRR; 6.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 8.
KW   Immunoglobulin domain.
FT   NON_TER       1       1
SQ   SEQUENCE   640 AA;  71949 MW;  6058974872636838 CRC64;


  Query Match             100.0%;  Score 3362;  DB 4;  Length 640;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-240;
  Matches  640;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MLNKMTLHPQQIMIGPRFNRALFDPLLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MLNKMTLHPQQIMIGPRFNRALFDPLLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKV 60

Qy      61 ICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAF 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAF 120

Qy     121 NGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 NGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLG 180

Qy     181 ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQ 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQ 240

Qy     241 GLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHH 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHH 300

Qy     301 NPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPP 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 NPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPP 360

Qy     361 ADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQD 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQD 420

Qy     421 TGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTP 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTP 480

Qy     481 VVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGCFVAITLMA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 VVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGCFVAITLMA 540

Qy     541 AVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPMPAIEHEHLNHYNS 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 AVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPMPAIEHEHLNHYNS 600

Qy     601 YKSPFNHTTTVNTINSIHSSVHEPLLIRMNSKDNVQETQI 640
           ||||||||||||||||||||||||||||||||||||||||
Db     601 YKSPFNHTTTVNTINSIHSSVHEPLLIRMNSKDNVQETQI 640


RESULT   2
Q9HBW1
ID   Q9HBW1      PRELIMINARY;      PRT;   653 AA.
AC   Q9HBW1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BRAIN TUMOR ASSOCIATED PROTEIN NAG14.
GN   NAG14.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wang J., Bin L., Jiang N., Li G.;
RT   "Homo sapiens brain-specific gene (BAG), downregulated in brain tumor,
RT   mRNA.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF196976; AAG28019.2; -.
DR   HSSP; P23945; 1XUN.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 8.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00370; LRR; 3.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 7.
KW   Immunoglobulin domain.
SQ   SEQUENCE   653 AA;  72717 MW;  38159C81F6850E37 CRC64;


  Query Match             55.5%;  Score 1865;  DB 4;  Length 653;
  Best Local Similarity   56.4%;  Pred. No. 1.2e-129;
  Matches  363;  Conservative  107;  Mismatches  130;  Indels   44;  Gaps   14;

Qy      29 VLLALQLLVVAGLVRA------QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLN 82
           | |  |: ::   : |      | ||||||||||||||:| |: | ||| || :||| ||
Db      22 VYLTAQVWILCAAIAAAASAGPQNCPSVCSCSNQFSKVVCTRRGLSEVPQGIPSNTRYLN 81

Qy      83 LHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNG 142
           | || ||:|: ::|:|| |||:||| || || ||:||||||:||||||||| || ||:|
Db      82 LMENNIQMIQADTFRHLHHLEVLQLGRNSIRQIEVGAFNGLASLNTLELFDNWLTVIPSG 141

Qy     143 AFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYL 202
           || |||||:|||||||||||||||||:||| ||||||||:| ||||||||| ||:||
Db     142 AFEYLSKLRELWLRNNPIESIPSYAFNRVPSLMRLDLGELKKLEYISEGAFEGLFNLKYL 201

Qy     203 NLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNA 262
           || |||:::|||||: |:||::|||   |||||| ||  |:|||:: ||: :|||||
Db     202 NLGMCNIKDMPNLTPLVGLEELEMSGNHFPEIRPGSFHGLSSLKKLWVMNSQVSLIERNA 261

Qy     263 FDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPS 322
           || | ||||:|||||||: |||||||| :|  :|||||||||:|||||:||:::  |:
Db     262 FDGLASLVELNLAHNNLSSLPHDLFTPLRYLVELHLHHNPWNCDCDILWLAWWLREYIPT 321

Qy     323 NTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLT 382
           |: || ||: | :::|||: |:||  | | || |:: | |||::||  ||||||  :  ::
Db     322 NSTCCGRCHAPMHMRGRYLVEVDQASFQCSAPFIMDAPRDLNISEGRMAELKCR-TPPMS 380

Qy     383 SVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLN 442
           || |: |||||::| :   ||:||:||||||::| : |||:|||||:|  ||: ||| ||
Db     381 SVKWLLPNGTVLSHASRHPRISVLNDGTLNFSHVLLSDTGVYTCMVTNVAGNSNASAYLN 440

Qy     443 VTAA--TTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTPVVDWETTNV------TTSLT 494
           |: |   |: :|:|:|||||| | | ::   |     | |     ||:       ||| |
Db     441 VSTAELNTSNYSFFTTVTVETTEISPED---TTRKYKPVP-----TTSTGYQPAYTTSTT 492

Qy     495 P--QSTRSTEKTFTIPVTDINSGI-PGIDEVMKTTKIIIGCFVAITLMAAVMLVIFYKMR 551
             |:||   |   :| ||    :   :|||||||||||||||:||:|| ||::|||:|
Db     493 VLIQTTR-VPKQVAVPATDTTDKMQTSLDEVMKTTKIIIGCFVAVTLLAAAMLIVFYKLR 551

Qy     552 KQHHRQNHHAPTRTVEIINVDDEITGDTPM----------ESHLPMPAIEHEHLNHYNS 600
           |:| :::     |||||| ||::|   |             |  : :| | |:|:| ||:
```

```
Db    552 KRHQQRSTVTAARTVEIIQVDEDIPAATSAAATAAPSGVSGEGAVVLPTI-HDHIN-YNT 609

Qy    601 YKSPFNHTTTVNTI-NSIH---SSVHEPLLIRMNSKDNVQETQI 640
          ||       | |:: ||:|   ::: || :|: ::|| ||||||
Db    610 YKPAHGAHWTENSLGNSLHPTVTTISEPYIIQTHTKDKVQETQI 653


RESULT   3
Q96A85
ID   Q96A85      PRELIMINARY;      PRT;   649 AA.
AC   Q96A85;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 72.3 KDA PROTEIN.
GN   BAG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wang J.;
RL   Thesis (2000), Zhongshan Medical University, Guangzhou, China.
DR   EMBL; AJ297858; CAC82651.1; -.
KW   Signal; Hypothetical protein.
FT   SIGNAL       1     38       POTENTIAL.
SQ   SEQUENCE   649 AA;  72313 MW;  9C11C5ABC7E536CF CRC64;


  Query Match             54.7%;  Score 1840;  DB 4;  Length 649;
  Best Local Similarity   56.1%;  Pred. No. 8.7e-128;
  Matches  361;  Conservative  105;  Mismatches  130;  Indels   48;  Gaps   15;

Qy     29 VLLALQLLVVAGLVRA------QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLN 82
          | |  |: ::   : |      | ||||||||||||||:| |: | ||| || :||| ||
Db     22 VYLTAQVWILCAAIAAAASAGPQNCPSVCSCSNQFSKVVCTRRGLSEVPQGIPSNTRYLN 81

Qy     83 LHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNG 142
          | || ||:|: ::|:|| |||:||| || || ||:|||||||:||||||| || ||:|
Db     82 LMENNIQMIQADTFRHLHHLEVLQLGRNSIRQIEVGAFNGLASLNTLELFDNWLTVIPSG 141

Qy    143 AFVYLSKLKELWLRNNPIESISPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYL 202
          || |||||:|||||||||||||||||:||| ||||||||:| |||||||||| ||:||
Db    142 AFEYLSKLRELWLRNNPIESIPSYAFNRVPSLMRLDLGELKKLEYISEGAFEGLFNLKYL 201

Qy    203 NLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNA 262
          || |||:::||||||: |:||::|||   |||||| ||  |:|||:: |     ||||
Db    202 NLGMCNIKDMPNLTPLVGLEELEMSGNHFPEIRPGSFHGLSSLKKLWVMNSH----ERNA 257

Qy    263 FDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPS 322
          || | ||||:|||||||: |||||||||| :|  :|||||||||:||||||:||:::  |:
Db    258 FDGLASLVELNLAHNNLSSLPHDLFTPLRYLVELHLHHNPWNCDCDILWLAWWLREYIPT 317

Qy    323 NTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLT 382
          |: || ||: | :::||: |:||  | | || |:: | |||::||  |||||| :  ::
Db    318 NSTCCGRCHAPMHMRGRYLVEVDQASFQCSAPFIMDAPRDLNISEGRMAELKCR-TPPMS 376

Qy    383 SVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLN 442
          || |: |||||::| :   ||:||:||||||::| : |||:|||||:|  ||: ||| ||
Db    377 SVKWLLPNGTVLSHASRHPRISVLNDGTLNFSHVLLSDTGVYTCMVTNVAGNSNASAYLN 436

Qy    443 VTAA--TTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTPVVDWETTNV------TTSLT 494
          |: |   |: :|:|:|||||| | | ::   |     | |    ||:       ||| |
Db    437 VSTAELNTSNYSFFTTVTVETTEISPED---TTRKYKPVP-----TTSTGYQPAYTTSTT 488

Qy    495 P--QSTRSTEKTFTIPVTDINSGI-PGIDEVMKTTKIIIGCFVAITLMAAVMLVIFYKMR 551
            |:||   |   :| ||    :   :|||||||||||||||:||:|| ||::|||:|
Db    489 VLIQTTR-VPKQVAVPATDTTDKMQTSLDEVMKTTKIIIGCFVAVTLLAAAMLIVFYKLR 547

Qy    552 KQHHRQNHHAPTRTVEIINVDDEITGDTPM----------ESHLPMPAIEHEHLNHYNS 600
          |:| :::     |||||| ||::|   |             |  : :| | |:|:| ||:
Db    548 KRHQQRSTVTAARTVEIIQVDEDIPAATSAAATAAPSGVSGEGAVVLPTI-HDHIN-YNT 605
```

```
Qy    601 YKSPFNHTTTVNTI-NSIH---SSVHEPLLIRMNSKDNVQETQI 640
          ||       | |:: ||:|   ::: || :|: ::|| ||||||
Db    606 YKPAHGAHWTENSLGNSLHPTVTTISEPYIIQTHTKDKVQETQI 649


RESULT   4
Q9NT99
ID   Q9NT99      PRELIMINARY;      PRT;   422 AA.
AC   Q9NT99;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 45.1 KDA PROTEIN (FRAGMENT).
GN   DKFZP761A179.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=AMYGDALA;
RA   Blum H., Bauersachs S., Mewes H.W., Gassenhuber J., Wiemann S.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL137451; CAB70743.1; -.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   InterPro; IPR000217; Tubulin.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 3.
DR   Pfam; PF01463; LRRCT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00370; LRR; 1.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00369; LRR_TYP; 2.
DR   PROSITE; PS00227; TUBULIN; UNKNOWN_1.
KW   Hypothetical protein; Immunoglobulin domain.
FT   NON_TER       1       1
FT   NON_TER     422     422
SQ   SEQUENCE   422 AA;  45110 MW;  164C6AF1A4BC440C CRC64;


  Query Match             40.9%;  Score 1376.5;  DB 4;  Length 422;
  Best Local Similarity   62.6%;  Pred. No. 9.2e-94;
  Matches  256;  Conservative   56;  Mismatches   58;  Indels   39;  Gaps    6;

Qy    206 MCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDN 265
          ||||::||||| |::|:||:|||| |  |||||||||  |:|||:: :|:  |||||||:
Db      1 MCNLKDIPNLTALVRLEELELSGNRLDLIRPGSFQGLTSLRKLWLMHAQVATIERNAFDD 60

Qy    266 LQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTA 325
          |:|| |:||:|||| ||||||||||| |||:||:||||:||||:||||||:|:  ||||
Db     61 LKSLEELNLSHNNLMSLPHDLFTPLHRLERVHLNHNPWHCNCDVLWLSWWLKETVPSNTT 120

Qy    326 CCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVS 385
          |||||: |  |||||||||||::||||||||||| ||||||||||||||  ||:|||:
Db    121 CCARCHAPAGLKGRYIGELDQSHFTCYAPVIVEPPTDLNVTEGMAAELKCRTGTSMTSVN 180

Qy    386 WITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTA 445
          |:|||||:||||:|:|||:|| |||||||||||||| |||||:|| ||||||||||:|
Db    181 WLTPNGTLMTHGSYRVRISVLHDGTLNFTNVTVQDTGQYTCMVTNSAGNTTASATLNVSA 240

Qy    446 ATTTP----------------------FSYFSTVTVETME--PSQD--EARTTDNN-VG 477
                                       ::||:||||||:|  | ::  : | |:    |
Db    241 VDPVAAGGTGSGGGGPGGSGGVGGGSGGYTYFTTVTVETLETQPGEEALQPRGTEKEPPG 300

Qy    478 PTPVVDW----------ETTNVTTSLTPQSTRSTEKTFTIPVTDI-NSGIPGIDEVMKTT 526
          ||    |            :: ||:  |:|:| ||| ||:|:||:  : :  :|:|||||
```

```
Db    301 PTTDGVWGGGRPGDAAGPASSSTTAPAPRSSRPTEKAFTVPITDVTENALKDLDDVMKTT 360

Qy    527 KIIIGCFVAITLMAAVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDDEI 575
          |||||||||||| ||||||| |||:|||| || ||||||||||:||:
Db    361 KIIIGCFVAITFMAAVMLVAFYKLRKQHQLHKHHGPTRTVEIINVEDEL 409


RESULT  5
Q99PH1
ID   Q99PH1      PRELIMINARY;      PRT;   441 AA.
AC   Q99PH1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LIBG-LIKE PROTEIN (FRAGMENT).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wang J., Bin L., Li G.;
RT   "Brain tumor-associated mRNA.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF290542; AAG60620.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 2.
DR   Pfam; PF01463; LRRCT; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00370; LRR; 2.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00369; LRR_TYP; 3.
KW   Immunoglobulin domain.
FT   NON_TER       1      1
SQ   SEQUENCE   441 AA;  48844 MW;  E2D17AB4607AC475 CRC64;


  Query Match             34.3%;  Score 1153;  DB 11;  Length 441;
  Best Local Similarity   50.3%;  Pred. No. 3.3e-77;
  Matches  224;  Conservative   83;  Mismatches  116;  Indels   22;  Gaps    9;

Qy    214 NLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEIN 273
          |||||: |:||::|||   |||||| ||  |:|||:: ||: :||||||| | ||||:|
Db      1 NLTPLVGLEELEMSGNHFPEIRPGSFHGLSSLKKLWVMNSQVSLIERNAFDGLASLVELN 60

Qy    274 LAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTP 333
          ||||||: |||||||||| :|  :|||||||||:|||||:||:::  |:|: || ||: |
Db     61 LAHNNLSSLPHDLFTPLRYLVELHLHHNPWNCDCDILWLAWWLREYIPTNSTCCGRCHAP 120

Qy    334 PNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTV 393
           :::|||: |:||  | | || |:: | |||::|   |||||| :  ::|| |: |||||
Db    121 MHMRGRYLVEVDQAAFQCSAPFIMDAPRDLNISEDRMAELKCR-TPPMSSVKWLLPNGTV 179

Qy    394 MTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAA--TTTPF 451
          ::| :   ||:||:|||||| | : |||:|||||:|  ||: ||| |||::|   |  |
Db    180 LSHASRHPRISVLNDGTLNFSRVLLIDTGVYTCMVTNVAGNSNASAYLNVSSAELNTPNF 239

Qy    452 SYFSTVTVETMEPSQDEARTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTD 511
          |:|:|||||| | | ::  |      ||    ::    |::     |    |   :| ||
Db    240 SFFTTVTVETTEISPEDI-TRKYKPVPTTSTGYQPAYTTSTTVLIQTTRVPKQVPVPSTD 298

Qy    512 INSGI-PGIDEVMKTTKIIIGCFVAITLMAAVMLVIFYKMRKQHHRQNHHAPTRTVEIIN 570
              :   :|||||||||||||||:||:|| ||::|||:||:| :::     ||||||
```

```
Db    299 TTDKMQTSLDEVMKTTKIIIGCFVAVTLLAAAMLIVFYKLRKRHQQRSTVTAARTVEIIQ 358

Qy    571 VDDEITGDTPM----------ESHLPMPAIEHEHLNHYNSYKSPFNHTTTVNTI-NSIH 618
          ||::|    |           |  : :| | |:|:| ||:||       | |:: ||:|
Db    359 VDEDIPAAAPAAATAAPSGVSGEGAVVLPTI-HDHIN-YNTYKPAHGAHWTENSLGNSLH 416

Qy    619 ---SSVHEPLLIRMNSKDNVQETQI 640
             ::: || :|: ::|| ||||||
Db    417 PTVTTISEPYIIQTHTKDKVQETQI 441


RESULT  6
Q96FE5
ID   Q96FE5      PRELIMINARY;      PRT;   614 AA.
AC   Q96FE5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   UNKNOWN (PROTEIN FOR MGC:17422).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN, AND ANAPLASTIC OLIGODENDROGLIOMA WITH 1P/19Q LOSS;
RA   Strausberg R.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC011057; AAH11057.1; -.
SQ   SEQUENCE   614 AA;  69145 MW;  EFD967E3B716698D CRC64;


  Query Match             17.4%;  Score 583.5;  DB 4;  Length 614;
  Best Local Similarity   27.7%;  Pred. No. 6.8e-35;
  Matches  182;  Conservative   96;  Mismatches  241;  Indels  137;  Gaps   16;

Qy     20 RALFDPLLLVVLLALQLLVVAGLV--RAQTCPSVCSCSNQFSKVICVRKNLREVPDGISTN 77
          |::  |||     : |||:  ::    |  ||  | || |   |:| ||    ||:|| |
Db      7 RSMPSPLLACWQPILLLVLGSVLSGSATGCPPRCECSAQDRAVLCHRKRFVAVPEGIPTE 66

Qy     78 TRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLT 137
          ||||:| :|:|: :  : |    ||| |:|: | :  :| |||| | || || |  |||
Db     67 TRLLDLGKNRIKTLNQDEFASFPHLEELELNENIVSAVEPGAFNNLFNLRTLGLRSNRLK 126

Qy    138 TIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLS 197
          || | |  || | :| :  | |  :  | |  : :|: |::|:   | |||  || ||:
Db    127 LIPLGVFTGLSNLTKLDISENKIVILLDYMFQDLYNLKSLEVGD-NDLVYISHRAFSGLN 185

Qy    198 NLRYLNLAMCNLREIP---------------------------------------------- 213
          :| | |  |||  ||
Db    186 SLEQLTLEKCNLTSIPTEALSHLHGLIVLRLRHLNINAIRDYSFKRLYRLKVLEISHWPY 245

Qy    214 ----------------------NLTP--------LIKLDELDLSGNHLSAIRPGSFQGLM 243
                               |||         |: |  |:|| | :| |       |:
Db    246 LDTMTPNCLYGLNLTSLSITHCNLTAVPYLAVRHLVYLRFLNLSYNPISTIEGSMLHELL 305

Qy    244 HLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPW 303
          ||:: ::  |: |:|  ||  | |  :|:: | || |   :|  : :|| : |  ||
Db    306 RLQEIQLVGGQLAVVEPYAFRGLNYLRVLNVSGNQLTTLEESVFHSVGNLETLILDSNPL 365

Qy    304 NCNCDILWL---SWWIKDMAPSNTACCARCNTPPNLKGRYIGE----LDQNYFTCYAPVI 356
          |:| :||:    | :       |  ||  ::|:   :    |  |||||    |
Db    366 ACDCRLLWVFRRRWRLNFNRQQPT-----CATPEFVQGKEFKDFPDVLLPNYFTCRRARI 420

Qy    357 VEPPA-DLNVTEGMAAELKCRA-STSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFT 414
          :  |  : | ||   :  |||      :: |::|   :::      |: |  ||||
Db    421 RDRKAQQVFVDEGHTVQFVCRADGDPPPAILWLSPRKHLVS-AKSNGRLTVFPDGTLEVR 479

Qy    415 NVTVQDTGMYTCMVSNSVGNTTASATLNVTA-ATTTPFSYFSTVTVETMEPSQDEARTTD 473
          ||| | |:|: :|: || :  | |:| : :   |      |     : :| : || :|
Db    480 YAQVQDNGTYLCIAANAGGNDSMPAHLHVRSYSPDWPHQPNKTFAFISNQPGEGEANSTR 539

Qy    474 NNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGCF 533
```

```
Db    540 ATV-PFP-FDIKTLIIATTMGFIS---------------------------------F 562

Qy    534 VAITLMAAVMLVIFY--KMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPM 587
Db    563 LGVVLFCLVLLFLWSRGKGNTKHNIEIEYVPRKSDAGIS-----SADAPRKFNMKM 613


RESULT   7
Q9N008
ID   Q9N008      PRELIMINARY;      PRT;   614 AA.
AC   Q9N008;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 69.2 KDA PROTEIN.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CEREBELLUM CORTEX;
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB046639; BAB03557.1; -.
DR   HSSP; P23945; 1XUN.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 9.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00370; LRR; 5.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 1.
KW   Hypothetical protein; Immunoglobulin domain.
SQ   SEQUENCE   614 AA;  69187 MW;  BA6C8BC7C993BE9A CRC64;


  Query Match           17.2%;  Score 578.5;  DB 6;  Length 614;
  Best Local Similarity  27.6%;  Pred. No. 1.6e-34;
  Matches  181;  Conservative   96;  Mismatches  242;  Indels  137;  Gaps   16;

Qy     20 RALFDPLLVVLLALQLLVVAGLV--RAQTCPSVCSCSNQFSKVICVRKNLREVPDGISTN 77
Db      7 RSMPSPLLACWQPILLLVLGSVLSGSATGCPPRCECSAQDRAVLCHRKRFVAVPEGIPTE 66

Qy     78 TRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLT 137
Db     67 TRLLDLGKNRIKTLNQDEFASFPHLEELELNENIVSAVEPGAFNNLFNLRTLGLRSNRLK 126

Qy    138 TIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLS 197
Db    127 LIPLGVFTGLSNLTKLDISENKIVILLDYMFQDLYNLKSLEVGD-NDLVYISHRAFSGLN 185

Qy    198 NLRYLNLAMCNLREIP---------------------------------------------- 213
Db    186 SLEQLTLEKCNLTSIPTEALSHLHGLIVLRLRHLNINAIRDYSFKRLYRLKVLEISHWPY 245

Qy    214 ----------------------NLTP--------LIKLDELDLSGNHLSAIRPGSFQGLM 243
```

```
Db    246 LDTMTPNCLYGLNLTSLSITHCNLTAVPYLAVRHLVYLRFLNLSYNPISTIEGSMLHELL 305

Qy    244 HLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPW 303
Db    306 RLQEIQLVGGQLAMVEPYAFRGLNYLRVLNVSGNQLTTLEESVFHSVGNLETLILDSNPL 365

Qy    304 NCNCDILWL---SWWIKDMAPSNTACCARCNTPPNLKGRYIGE----LDQNYFTCYAPVI 356
Db    366 ACDCRLLWVFRRRWRLNFNRQQPT-----CATPEFVQGKEFKDFPDVLLPNYFTCRRARI 420

Qy    357 VEPPA-DLNVTEGMAAELKCRA-STSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFT 414
Db    421 RDRKAQQVFVDEGHTVQFVCRADGDPPPAILWLSPRKHLVS-AKSNGRLTVFPDGTLEVR 479

Qy    415 NVTVQDTGMYTCMVSNSVGNTTASATLNVTA-ATTTPFSYFSTVTVETMEPSQDEARTTD 473
Db    480 YAQVQDNGTYLCIAANAGGNDSMPAHLHVRSYSPDWPHQPNKTFAFIPNQPGEGEANSTR 539

Qy    474 NNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGCF 533
Db    540 ATV-PFP-FDIKTLIIATTMGFIS---------------------------------F 562

Qy    534 VAITLMAAVMLVIFY--KMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPM 587
Db    563 LGVVLFCLVLLFLWSRGKGNTKHNIEIEYVPRKSDAGIS-----SADAPRKFNMKM 613


RESULT   8
Q9D1T0
ID   Q9D1T0      PRELIMINARY;      PRT;   614 AA.
AC   Q9D1T0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ADULT MALE TESTIS CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
DE   CLONE:4930471K13, FULL INSERT SEQUENCE.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TESTIS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK027262; BAB32403.1; -.
DR   HSSP; P23945; 1XUN.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
```

```
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 9.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 9.
KW   Immunoglobulin domain.
SQ   SEQUENCE   614 AA;  69100 MW;  41CFF40C21335681 CRC64;


  Query Match           17.2%;  Score 577.5;  DB 11;  Length 614;
  Best Local Similarity   27.6%;  Pred. No. 1.9e-34;
  Matches  181;  Conservative   95;  Mismatches  243;  Indels  137;  Gaps   16;

Qy      20 RALFDPLLVVLLALQLLVVAGLV--RAQTCPSVCSCSNQFSKVICVRKNLREVPDGISTN 77
           |::  |||     : |||:  ::   |  ||  | || |   |:| ||    ||:|| |
Db       7 RSMPSPLLACWQPILLLVLGSVLSGSATGCPPRCECSAQDRAVLCHRKRFVAVPEGIPTE 66

Qy      78 TRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLT 137
           ||||:| :|:|: :  : |    ||| |:|: | :  :| |||| | || || |  |||
Db      67 TRLLDLGKNRIKTLNQDEFASFPHLEELELNENIVSAVEPGAFNNLFNLRTLGLRSNRLK 126

Qy     138 TIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLS 197
           || | |  || | :| :  | |  :  | |  : :|: |::|:   | |||  || ||:
Db     127 LIPLGVFTGLSNLTKLDISENKIVILLDYMFQDLYNLKSLEVGD-NDLVYISHRAFSGLN 185

Qy     198 NLRYLNLAMCNLREIP---------------------------------------------- 213
           :|  | |  |||  ||
Db     186 SLEQLTLEKCNLTSIPTEALSHLHGLIVLRLRHLNINAIRDYSFKRLYRLKVLEISHWPY 245

Qy     214 ----------------------NLTP--------LIKLDELDLSGNHLSAIRPGSFQGLM 243
                               |||         |: |  |:|| | : |        |:
Db     246 LDTMTPNCLYGLNLTSLSITHCNLTAVPYLAVRHLVYLRFLNLSYNPIGTIEGSMLHELL 305

Qy     244 HLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPW 303
           ||:: ::  |: |:|  ||  |  |  :|:: | || |     |  : :|| : |  ||
Db     306 RLQEIQLVGGQLAVVEPYAFRGLNYLRVLNVSGNQLTTLEESAFHSVGNLETLILDSNPL 365

Qy     304 NCNCDILWL---SWWIKDMAPSNTACCARCNTPPNLKGRYIGE----LDQNYFTCYAPVI 356
           |:| :||:    | :       |     | ||  ::|:   :     |  |||||    |
Db     366 ACDCRLLWVFRRRWRLNFNRQQPT-----CATPEFVQGKEFKDFPDVLLPNYFTCRRAHI 420

Qy     357 VEPPA-DLNVTEGMAAELKCRA-STSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFT 414
           :  |  : | ||   :  |||      :: |::|   :::      |: |  ||||
Db     421 RDRKAQQVFVDEGHTVQFVCRADGDPPPAILWLSPRKHLVS-AKSNGRLTVFPDGTLEVR 479

Qy     415 NVTVQDTGMYTCMVSNSVGNTTASATLNVTA-ATTTPFSYFSTVTVETMEPSQDEARTTD 473
             ||| | | |: :|: || :  | |:| : :   |     |    : :| : || :|
Db     480 YAQVQDNGTYLCIAANAGGNDSMPAHLHVRSYSPDWPHQPNKTFAFISNQPGEGEANSTR 539

Qy     474 NNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGCF 533
           | | |  | :|  : |::  |                                      |
Db     540 ATV-PFP-FDIKTLIIATTMGFIS----------------------------------F 562

Qy     534 VAITLMAAVMLVIFY--KMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPM 587
           : : |   |:| ::   |  :|: :  : | ::   |:     : | | : :: |
Db     563 LGVVLFCLVLLFLWSRGKGNTKHNIEIEYVPRKSDAGIS-----SADAPRKFNMKM 613


RESULT   9
Q9BZ20
ID   Q9BZ20      PRELIMINARY;      PRT;   606 AA.
AC   Q9BZ20;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BA438B23.1 (NEURONAL LEUCINE-RICH REPEAT PROTEIN) (CDNA FLJ31810 FIS,
DE   CLONE NT2RI2009289, WEAKLY SIMILAR TO CARBOXYPEPTIDASE N 83 KDA
DE   CHAIN).
GN   BA438B23.1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Babbage A.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA   Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
RA   Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA   Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA   Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA   Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL353746; CAC22713.1; -.
DR   EMBL; AK056372; BAB71167.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 11.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 10.
KW   Immunoglobulin domain.
SQ   SEQUENCE   606 AA;  68065 MW;  CB608E281B066B9D CRC64;


  Query Match           17.1%;  Score 575;  DB 4;  Length 606;
  Best Local Similarity   26.9%;  Pred. No. 2.8e-34;
  Matches  172;  Conservative   95;  Mismatches  224;  Indels  148;  Gaps   16;

Qy      31 LALQLLVVAGLVRAQTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQI 90
           ||: |: :   :    ||: | || |   | | |: |  :|:||   |::|:| :|::: 
Db      15 LAVVLIFMGSTI---GCPARCECSAQNKSVSCHRRRLIAIPEGIPIETKILDLSKNRLKS 71

Qy      91 IKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKL 150
           :    |     || : || | |  :| |||| | || :| |  |||  :| | |  || |
Db      72 VNPEEFISYPLLEEIDLSDNIIANVEPGAFNNLFNLRSLRLKGNRLKLVPLGVFTGLSNL 131

Qy     151 KELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLR 210
           :| :  | |  :  | |  : :|: |::|:   | |||  || || :|  | |  |||
Db     132 TKLDISENKIVILLDYMFQDLHNLKSLEVGD-NDLVYISHRAFSGLLSLEQLTLEKCNLT 190

Qy     211 EIP-------------------------------------------------------N 214
           :|                                                        |
Db     191 AVPTEALSHLRSLISLHLKHLNINNMPVYAFKRLFHLKHLEIDYWPLLDMMPANSLYGLN 250

Qy     215 LTP------------------LIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQ 256
           ||                   |: |  |:|| | :| |  | |  |: ||:| :: :|::
Db     251 LTSLSVTNTNLSTVPFLAFKHLVYLTHLNLSYNPISTIEAGMFSDLIRLQELHIVGAQLR 310

Qy     257 VIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWL---- 312
           || ::|  |: |  :|:: | |  |  ::|:    || : :::||  |:| :||:
```

```
Db    311 TIEPHSFQGLRFLRVLNVSQNLLETLEENVFSSPRALEVLSINNNPLACDCRLLWILQRQ 370

Qy    313 -SWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQN----YFTCYAPVIVEPPAD-LNVT 366
Db    371 PTLQFGGQQP-------MCAGPDTIRERSFKDFHSTALSFYFTCKKPKIREKKLQHLLVD 423

Qy    367 EGMAAELKCRASTSLTSV-SWITPNG---TVMTHGAYKVRIAVLSDGTLNFTNVTVQDTG 422
Db    424 EGQTVQLECSADGDPQPVISWVTPRRRFITTKSNG----RATVLGDGTLEIRFAQDQDSG 479

Qy    423 MYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTPVV 482
Db    480 MYVCIASNAAGNDTFTASLTVKGFASDRFLYANRTPMYMTDSNDTISNGTNAN---TFSL 536

Qy    483 DWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIGCFVAITLMAAV 542
Db    537 DLKTILVSTAM----------------------------------GCFTFLGVVLFC 559

Qy    543 MLVIFYKMR-KQHHRQN---HHAPTRTVEIINVDDEITG 577
Db    560 FLLLFVWSRGKGKHKNSIDLEYVPRKNNGAV-VEGEVAG 597


RESULT  10
Q96NI6
ID   Q96NI6      PRELIMINARY;      PRT;   719 AA.
AC   Q96NI6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CDNA FLJ30803 FIS, CLONE FEBRA2001245, WEAKLY SIMILAR TO NAG14.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA   Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA   Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA   Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK055365; BAB70910.1; -.
SQ   SEQUENCE   719 AA;  79472 MW;  CE301F219705FF4B CRC64;


  Query Match          14.3%;  Score 480.5;  DB 4;  Length 719;
  Best Local Similarity   25.5%;  Pred. No. 3.5e-27;
  Matches  169;  Conservative   78;  Mismatches  249;  Indels  167;  Gaps   16;

Qy     29 VLLALQLLVVAGLVRAQTCPSVCSCS--NQFSKVICVRKNLREVPDGISTNTRLLNLHEN 86
Db      4 ILFYLFLIGIA--VKAQICPKRCVCQILSPNLATLCAKKGLLFVPPNIDRRTVELRLADN 61

Qy     87 QIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVY 146
Db     62 FVTNIKRKDFANMTSLVDLTLSRNTISFITPHAFADLRNLRALHLNSNRLTKITNDMFSG 121

Qy    147 LSKLKELWLRNNPIESISPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM 206
Db    122 LSNLHHLILNNNQLTLISSTAFDDV----------------------------------146

Qy    207 CNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNL 266
Db    147 -----------FALEELDLSYNNLETIPWDAVEKMVSLHTLSLDHNMIDNIPKGTFSHL 194

Qy    267 QSLVEINLAHNNLTLLPHDLFTPLHHLERI-------------HLHHNPWNCNCDILWL 312

          :  :::  | |  || |   ||    ::                    || :|||::|||
Db    195 HKMTRLDVTSNKLQKLPPD---PLFQRAQVLATSGIISPSTFALSFGGNPLHCNCELLWL 251

Qy    313 SWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMAAE 372
Db    252 ------RRLSREDDLETCASPPLLTGRYFWSIPEEEFLCEPPLITRHTHEMRVLEGQRAT 305

Qy    373 LKCRA-STSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNS 431
Db    306 LRCKARGDPEPAIHWISPEGKLISNA---TRSLVYDNGTLDILITTVKDTGAFTCIASNP 362

Qy    432 VGNTTASA---------TLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTPVV 482
Db    363 AGEATQIVDLHIIKLPHLLNSTNHIHEPDPGSSDISTSTKSGSNTSSSNGDTKLSQDKIV 422

Qy    483 DWETTNVTTSLTPQSTRS----------------------------TEKTFTIPVTDINSGI 516
Db    423 VAEATSSTALLKFNFQRNIPGIRMFQIQYNGTYDDTLVYRMIPPTSKTFL--VNNLAAGT 480

Qy    517 -----------PGIDEVMKT----------------------------TKIII--GCFVAI 536
Db    481 MYDLCVLAIYDDGITSLTATRVVGCIQFTTEQDYVRCHFMQSQFLGGTMIIIIGGIIVAS 540

Qy    537 TLMAAVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPMP------AI 590
Db    541 VLVFIIILMIRYKV------CNNNGQHKVTKVSNVYSQTNGAQIQGCSVTLPQSVSKQAV 594

Qy    591 EHE 593
Db    595 GHE 597


RESULT  11
Q9V430
ID   Q9V430      PRELIMINARY;      PRT;  1021 AA.
AC   Q9V430;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KEK3 PROTEIN.
GN   KEK3 OR BG:DS04862.1 OR CG4192.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
```

```
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Y, AND CN BW SP;
RX   MEDLINE=99403001; PubMed=10471707;
RA   Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,
RA   Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA   Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA   Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
RA   Celniker S., Rubin G.M.;
RT   "An exploration of the sequence of a 2.9-Mb region of the genome of
RT   Drosophila melanogaster: the Adh region.";
RL   Genetics 153:179-219(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Y, AND CN BW SP;
RA   Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA   Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA   Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA   Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA   Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA   Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA   Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
RA   Zieran L.L., Rubin G.M.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003648; AAF53467.1; -.
DR   EMBL; AE003413; AAF44948.1; -.
DR   FlyBase; FBgn0028370; kek3.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 6.
DR   Pfam; PF01463; LRRCT; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00370; LRR; 2.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 2.
KW   Immunoglobulin domain.
SQ   SEQUENCE   1021 AA;  109770 MW;  23CB5712EFD60969 CRC64;


  Query Match             13.9%;  Score 466;  DB 5;  Length 1021;
  Best Local Similarity   30.6%;  Pred. No. 6.6e-26;
  Matches  142;  Conservative   57;  Mismatches  157;  Indels  108;  Gaps   16;

Qy      15 GPRFNRALFDPL-----LVVLLALQLLVVA---GLVRAQTCPSVCSCSNQFSK--VICVR 64
           |||  | |  ||     | : | | ||      | :||: ||:|| |  :  |  |:|:
Db      40 GPRPQRRLHPPLRPRLPLHLHLLLWLLCCCSQLGQLRAE-CPAVCECKWKSGKESVLCLN 98

Qy      65 KNLREVPDGISTNTRLLNLHENQIQIIKVNSF--KHLRHLEILQLSRNHIRTIEIGAFNG 122
             ||  :|  :   |:||:|  |:||:|  :||    | :|: : |:| |:| ||  ||
Db      99 ANLTHIPQPLDAGTQLLDLSGNEIQLIPDDSFATAQLLNLQKVYLARCHLRLIERHAFRK 158

Qy     123 LANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGEL 182
           | ||  |:|  | |: ||: |  ::|:|:|| |  |||  :|  ||  :| | :|:| :
Db     159 LINLVELDLSQNLLSAIPSLALYHVSELRELRLSGNPILRVPDDAFGHVPQLVKLELSDC 218

Qy     183 KRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQGL 242
            |||:|:  || || :                    |: | | || || :| |:   |
Db     219 -RLSHIAVRAFAGLES--------------------SLEWLKLDGNRLSEVRSGTITSL 256

Qy     243 MHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNP 302
                                                  || ||   |  |
Db     257 ----------------------------------------------ASLHGLE---LARNT 268

Qy     303 WNCNCDILWL-SWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPA 361
           |||:| :  | :| ::   ||       | :|| | ||   :|| : | |  | ||
Db     269 WNCSCSLRPLRAWMLQQNIPSGIP--PTCESPPRLSGRAWDKLDVDDFAC-VPQIVATDT 325

Qy     362 DLNVTEGMAAELKCRA-STSLTSVSWITPNGTVMTHGA-------YKVRIAVLSDG---- 409
             :  ||    : |        :| |:  |  :    |        : | |  : |
Db     326 TAHGVEGRNITMSCYVEGVPQPAVKWLLKNRLIANLSAGGDGDSDSEPRTAAATQGRKTY 385

Qy     410 ----TLNFTNVTV-----QDTGMYTCMVSNSVGNTTASATLNVT 444
                 | :|:|:     || |:|||   |  |   || || |:
Db     386 VVNMLRNASNLTILTADMQDAGIYTCAAENKAGKVEASVTLAVS 429


RESULT  12
Q9ULH4
ID   Q9ULH4      PRELIMINARY;      PRT;   832 AA.
AC   Q9ULH4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA1246 PROTEIN (FRAGMENT).
GN   KIAA1246.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=20039619; PubMed=10574462;
RA   Nagase T., Ishikawa K., Kikuno R., Hirosawa M., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XV.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:337-345(1999).
DR   EMBL; AB033072; BAA86560.1; -.
DR   HSSP; P56276; 1TLK.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00041; fn3; 1.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 7.
DR   Pfam; PF01463; LRRCT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00060; FN3; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00370; LRR; 1.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 3.
KW   Immunoglobulin domain.
FT   NON_TER       1      1
SQ   SEQUENCE   832 AA;  88978 MW;  9B0E80F7BFF974A8 CRC64;
```

```
  Query Match           13.2%;  Score 442.5;  DB 4;  Length 832;
  Best Local Similarity  24.5%;  Pred. No. 2.7e-24;
  Matches  164;  Conservative   87;  Mismatches  245;  Indels  173;  Gaps   18;

Qy      20 RALFDPLLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKV--ICVRKNLREVPDGISTN 77
           :|| |  :  ||   |           ||  | | |    :  :|  | |  ||  |
Db      37 QALSDQTMETLLGGLLAFGMAFAVVDACPKYCVCQNLSESLGTLCPSKGLLFVPPDIDRR 96

Qy      78 TRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLT 137
           |  | |  | |    | ::  |  | |||| |  |:  :|  | :| :| |  |||
Db      97 TVELRLGGNFIIHISRQDFANMTGLVDLTLSRNTISHIQPFSFLDLESLRSLHLDSNRLP 156

Qy     138 TIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLS 197
           ::       |  |: | : ||                          :|  |:: |||
Db     157 SLGEDTLRGLVNLQHLIVNNN------------------------QLGGIADEAFEDF- 190

Qy     198 NLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQV 257
                           |: |::|||| |:|  :   | : :::| :| :  : :
Db     191 --------------------LLTLEDLDLSYNNLHGLPWDSVRRMVNLHQLSLDHNLLDH 230

Qy     258 IERNAFDNLQSLVEINLAHNNLTLLPHD-LF----------TPLHHLERIHLHHNPWNCN 306
           |    | :|| |  ::|  | |  || | :|          ||           || :||
Db     231 IAEGTFADLQKLARLDLTSNRLQKLPPDPIFARSQASALTATPFAPPLSFSFGGNPLHCN 290

Qy     307 CDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVT 366
           |::|||    :|          | :|  |||||   : :  | |  |:| :    | |
Db     291 CELLWLRRLERD------DDLETCGSPGGLKGRYFWHVREEEFVCEPPLITQHTHKLLVL 344

Qy     367 EGMAAELKCRA-STSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYT 425
           || || |||:|       : |: |:  :: : :   | ||  :|||:    | ||:| :|
Db     345 EGQAATLKCKAIGDPSPLIHWVAPDDRLVGNSS---RTAVYDNGTLDIFITTSQDSGAFT 401

Qy     426 CMVSNSVGNTTASATL---------NVTAATTTPFSYFSTVT------------------ 458
           |: :|: |  ||   :         | |: |  | |  | :|
Db     402 CIAANAAGEATAMVEVSIVQLPHLSNSTSRTAPPKSRLSDITGSSKTSRGGGGSGGGEPP 461

Qy     459 ----------------------------------------------VETMEPSQDEARTT 472
                                                        :  | |: ::|
Db     462 KSPPERAVLVSEVTTTSALVKWSVSKSAPRVKMYQLQYNCSDDEVLIYRMIPASNKAFVV 521

Qy     473 DNNVGPT-----PVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGI----DEVM 523
           :| |  |       :  |: |   |:||  :     : ||       :  |       :::
Db     522 NNLVSGTGYDLCVLAMWDDT--ATTLTATNIVGCAQFFT------KADYPQCQSMHSQIL 573

Qy     524 KTTKIII--GCFVAITLMAAVMLVIFYKMRKQHHRQNHHAPTR-TVEIINVDDEITGDTP 580
             | |::  |  ||  |:  |:|:: ||:       || ||::    : ||  :  |  |
Db     574 GGTMILVIGGIIVATLLVFIVILMVRYKV------CNHEAPSKMAAAVSNVYSQTNGAQP 627

Qy     581 MESHLPMPA 589
               | |:
Db     628 -----PPPS 631


RESULT  13
Q9BE71
ID   Q9BE71      PRELIMINARY;      PRT;   789 AA.
AC   Q9BE71;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 84.7 KDA PROTEIN.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FRONTAL LOBE LEFT;
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB056799; BAB39323.1; -.
DR   HSSP; P56276; 1TLK.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00041; fn3; 1.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 7.
DR   Pfam; PF01463; LRRCT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00060; FN3; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00370; LRR; 5.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 6.
KW   Hypothetical protein; Immunoglobulin domain.
SQ   SEQUENCE   789 AA;  84730 MW;  BB86DE81BC284B23 CRC64;


  Query Match           13.0%;  Score 437.5;  DB 6;  Length 789;
  Best Local Similarity  24.5%;  Pred. No. 6e-24;
  Matches  163;  Conservative   88;  Mismatches  237;  Indels  177;  Gaps   19;

Qy      24 DPLLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKV--ICVRKNLREVPDGISTNTRLL 81
           : ||  |||  :           ||  | | |    :  :|  | |  ||  |    |  |
Db       2 ETLLGGLLAFGMAFAV----VDACPKYCVCQNLSESLGTLCPSKGLLFVPPDIDRRTVEL 57

Qy      82 NLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPN 141
           |  | |  |    | ::  |  | |||| |  |:  :|  | :| :| |  |||  ::
Db      58 RLGGNFIIHISRQDFANMTGLVDLTLSRNTISHIQPFSFLDLESLRSLHLDSNRLPSLGE 117

Qy     142 GAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRY 201
                 |  |: | : ||                          :|  |:: |||
Db     118 DTLRGLVNLQHLIVNNN------------------------QLGGIADEAFEDF----- 147

Qy     202 LNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERN 261
                           |: |::|||| |:|  :   | : :::| :| :  : :  |
Db     148 ---------------LLTLEDLDLSYNNLHGLPWDSVRRMVNLHQLSLDHNLLDHIAEG 191

Qy     262 AFDNLQSLVEINLAHNNLTLLPHD-LF----------TPLHHLERIHLHHNPWNCNCDIL 310
           | :|| |  ::|  | |  || | :|          ||           || :|||::|
Db     192 TFADLQKLARLDLTSNRLQKLPPDPIFARSQASALTATPFAPPLSFSFGGNPLHCNCELL 251

Qy     311 WLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMA 370
           ||    :|           | :|  ||||   : :  | |  |:| :    | | || |
Db     252 WLRRLERD------DDLETCGSPGGLKGRYFWHVREEEFVCEPPLITQHTHKLLVLEGQA 305

Qy     371 AELKCRA-STSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVS 429
           | |||:|       : |: |:  :: : :   | ||  :|||:    | ||:| :||: :
Db     306 ATLKCKAIGDPSPLIHWVAPDDRLVGNSS---RTAVYDNGTLDIFITTSQDSGAFTCIAA 362

Qy     430 NSVGNTTASATL---------NVTAATTTPFSYFSTVT---------------------- 458
           |: |  ||:  :          | |: |  | |  | :|
Db     363 NAAGEATATVEVSIVQLPHLSNSTSRTAPPKSRLSDITGSSKTSRGGGGSGGGEPPKSPP 422

Qy     459 ----------------------------------------VETMEPSQDEARTTDNNV 476
                                                    :  | |: ::|    :| |
Db     423 ERAVLVSEVTTTSALAKWSVSKSTPRVKMYQLQYNCSDDEVLIYRMIPASNKAFVVNNLV 482

Qy     477 GPT-----PVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGI----DEVMKTTK 527
             |       :  |: |   |:||  :     : ||       :  |       :::  |
Db     483 SGTGYDLCVLAMWDDT--ATTLTATNIVGCAQFFT------KADYPQCQSMHSQILGGTM 534

Qy     528 III--GCFVAITLMAAVMLVIFYKMRKQHHRQNHHAPTR-TVEIINVDDEITGDTPMESH 584
```

```
          |::  |  ||  |:  |:|:: ||:       || ||::    : ||  :  |  |
Db    535 ILVIGGIIVATLLVFIVILMVRYKV------CNHEAPSKMAAAVSNVYSQTNGAQP---- 584

Qy    585 LPMPA 589
          | |:
Db    585 -PPPS 588


RESULT  14
O43377
ID   O43377      PRELIMINARY;      PRT;   705 AA.
AC   O43377;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PROBABLE LEUCINE-RICH REPEAT PROTEIN.
GN   RG118D07.1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FIBROBLAST;
RA   Murray J., Langston Y., Clarke C.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC004142; AAC02752.1; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00041; fn3; 1.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 9.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00370; LRR; 4.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   705 AA;  79049 MW;  9B46066693F4BC1A CRC64;


  Query Match           12.8%;  Score 432;  DB 4;  Length 705;
  Best Local Similarity   24.0%;  Pred. No. 1.3e-23;
  Matches  166;  Conservative  104;  Mismatches  250;  Indels  172;  Gaps   21;

Qy      25 PLLV-VLLALQL-LVVAGLVRAQTCPSVCSCS----------NQFSKVICVRKNLREVP 71
           || : ||| | :  :|  : :   || :|:|           : | | |    |   |
Db       2 PLRIHVLLGLAITTLVQAVDKKVDCPRLCTCEIRPWFTPRSIYMEASTVDCNDLGLLTFP 61

Qy      72 DGISTNTRLLNLHENQIQIIK------VN----------------SFKHLRHLEILQLS 108
             :  ||::| |  | |  |:      ||                : | :  |  : |
Db      62 ARLPANTQILLLQTNNIAKIEYSTDFPVNLTGLDLSQNNLSSVTNINVKKMPQLLSVYLE 121

Qy     109 RNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAF 168
           | :  :     : |:||  | :  | |:||  |||: |  |  | | :| :: | |  |
Db     122 ENKLTELPEKCLSELSNLQELYINHNLLSTISPGAFIGLHNLLRLHLNSNRLQMINSKWF 181

Qy     169 NRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPN-------------- 214
           : :|:|  | :||   :  | :  |: | ||| | :|  || |||:
Db     182 DALPNLEILMIGE-NPIIRIKDMNFKPLINLRSLVIAGINLTEIPDNALVGLENLESISF 240

Qy     215 -----------LTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWM--------IQS- 253
                       |  :: |  |||: | :: || | |  ::||::| :        | |
Db     241 YDNRLIKVPHVALQKVVNLKFLDLNKNPINRIRRGDFSNMLHLKELGINNMPELISIDSL 300

Qy     254 ----------------QIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERI 296
                            ::  |  |||  |  |  : |  | |: | |     | :|: |
Db     301 AVDNLPDLRKIEATNNPRLSYIHPNAFFRLPKLESLMLNSNALSALYHGTIESLPNLKEI 360

Qy     297 HLHHNPWNCNCDILWLSW---WIKDMAPSNTACCARCNTPPNLKGRYIGELD-QNYFTCY 352
           :| ||  |:| | |::     |: | | :  |      ||  :|: : ::  ::
Db     361 SIHSNPIRCDCVIRWMNMNKTNIRFMEPDSLFCV----DPPEFQGQNVRQVHFRDMMEIC 416

Qy     353 APVIVEP--PADLNVTEGMAAELKCRASTS-LTSVSWITPNGTVMTHGAYKVRIAVLSDG 409
           |:|     |::|||  |      |||:      : ||||:|  :       :  | |:|
Db     417 LPLIAPESFPSNLNVEAGSYVSFHCRATAEPQPEIYWITPSGQKLLPNTLTDKFYVHSEG 476

Qy     410 TLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEA 469
           ||:   || :: |:|||: :| ||    |  : |                    | :
Db     477 TLDINGVTPKEGGLYTCIATNLVGADLKSVMIKV----------------------DGS 513

Qy     470 RTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKII 529
           |||                          :  | : || :     : |: : |
Db     514 FPQDNN-------------------------GSLNIKIRDIQA-----NSVLVSWKA- 540

Qy     530 IGCFVAITLMAAVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPMPA 589
           :  | ::|    | |    |  |:   |:  |:: |:   :   |  :  : :|
Db     541 ----SSKILKSSVKWTAFVKTENSHAAQSARIPS-DVKVYNL-THLNPSTEYKICIDIPT 594

Qy     590 IEHEHLN----------HYNSYKSPFNHTTTV 611
           |  ::            | :  :   |:|||:
Db     595 IYQKNRKKCVNVTTKGLHPDQKEYEKNNTTTL 626


RESULT  15
Q9H3W5
ID   Q9H3W5      PRELIMINARY;      PRT;   708 AA.
AC   Q9H3W5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 79.4 KDA PROTEIN (NEURONAL LEUCINE-RICH REPEAT
DE   PROTEIN-3).
GN   DKFZP761K2424 OR NLRR-3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=AMYGDALA;
RA   Bloecker H., Boecher M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Hamano S., Inuzuka H., Morohashi A., Ohira M., Nakagawara A.;
RT   "Human neuronal leucine-rich repeat protein-3(NLRR-3).";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL442092; CAC09450.1; -.
DR   EMBL; AB060967; BAB47184.1; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00041; fn3; 1.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 9.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00060; FN3; 1.
DR   SMART; SM00409; IG; 1.
```

```
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00370; LRR; 4.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 8.
KW   Hypothetical protein; Immunoglobulin domain.
SQ   SEQUENCE   708 AA;  79424 MW;  24710478D6B124D1 CRC64;


  Query Match             12.8%;  Score 432;  DB 4;  Length 708;
  Best Local Similarity   24.0%;  Pred. No. 1.3e-23;
  Matches  166;  Conservative  104;  Mismatches  250;  Indels  172;  Gaps   21;

Qy      25 PLLV-VLLALQL-LVVAGLVRAQTCPSVCSCS-----------NQFSKVICVRKNLREVP 71
           || : ||| | :  :|  : :   || :|:|             : | | |    |   |
Db       5 PLRIHVLLGLAITTLVQAVDKKVDCPRLCTCEIRPWFTPRSIYMEASTVDCNDLGLLTFP 64

Qy      72 DGISTNTRLLNLHENQIQIIK------VN----------------SFKHLRHLEILQLS 108
             :  ||::| |  | |  |:      ||                : | :  |  : |
Db      65 ARLPANTQILLLQTNNIAKIEYSTDFPVNLTGLDLSQNNLSSVTNINVKKMPQLLSVYLE 124

Qy     109 RNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAF 168
            | :  :     : |:||  | :  | |:||  |||: |  |  | | :| :: | |  |
Db     125 ENKLTELPEKCLSELSNLQELYINHNLLSTISPGAFIGLHNLLRLHLNSNRLQMINSKWF 184

Qy     169 NRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPN-------------- 214
           : :|:|  | :||   :  | :  |: | ||| | :|  || |||:
Db     185 DALPNLEILMIGE-NPIIRIKDMNFKPLINLRSLVIAGINLTEIPDNALVGLENLESISF 243

Qy     215 -----------LTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWM--------IQS- 253
                       |  :: |  |||: | :: || | |  ::||::| :         | |
Db     244 YDNRLIKVPHVALQKVVNLKFLDLNKNPINRIRRGDFSNMLHLKELGINNMPELISIDSL 303

Qy     254 ----------------QIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERI 296
                          ::  |  |||  |  |  : |  | |: | |     | :|: |
Db     304 AVDNLPDLRKIEATNNPRLSYIHPNAFFRLPKLESLMLNSNALSALYHGTIESLPNLKEI 363

Qy     297 HLHHNPWNCNCDILWLSW---WIKDMAPSNTACCARCNTPPNLKGRYIGELD-QNYFTCY 352
           :| ||  |:| | |::     |: | | :  |      ||  :|: : ::  ::
Db     364 SIHSNPIRCDCVIRWMNMNKTNIRFMEPDSLFCV----DPPEFQGQNVRQVHFRDMMEIC 419

Qy     353 APVIVEP--PADLNVTEGMAAELKCRASTS-LTSVSWITPNGTVMTHGAYKVRIAVLSDG 409
            |:|     |::|||  |      |||:      : ||||:|  :       :  | |:|
Db     420 LPLIAPESFPSNLNVEAGSYVSFHCRATAEPQPEIYWITPSGQKLLPNTLTDKFYVHSEG 479

Qy     410 TLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEA 469
           ||:   || :: |:|||: :| ||    |  : |                    | :
Db     480 TLDINGVTPKEGGLYTCIATNLVGADLKSVMIKV----------------------DGS 516

Qy     470 RTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKII 529
             |||                          :  | : || :     : |: : |
Db     517 FPQDNN--------------------------GSLNIKIRDIQA-----NSVLVSWKA- 543

Qy     530 IGCFVAITLMAAVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPMPA 589
                :  | ::|    | |    |  |:   |:  |:: |:   :   |  :  : :|
Db     544 ----SSKILKSSVKWTAFVKTENSHAAQSARIPS-DVKVYNL-THLNPSTEYKICIDIPT 597

Qy     590 IEHEHLN----------HYNSYKSPFNHTTTV 611
           |  ::           | :  :   |:|||:
Db     598 IYQKNRKKCVNVTTKGLHPDQKEYEKNNTTTL 629


Search completed: August 26, 2002, 15:45:16
Job time: 253 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 2002, 15:37:37 ; Search time 14.64 Seconds
(without alignments)
1067.786 Million cell updates/sec

Title: US-09-905-056-292
Perfect score: 3362
Sequence: 1 MLNKMTLHPQQIMIGPRFNR.....VHEPLLIRMNSKDNVQETQI 640

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	432	12.8	708	4	US-09-131-648-2	Sequence 2, Appli
2	417.5	12.4	1091	3	US-08-986-485-5	Sequence 5, Appli
3	401	11.9	1101	3	US-08-986-485-2	Sequence 2, Appli
4	395.5	11.8	1525	3	US-09-191-647-2	Sequence 2, Appli
5	395.5	11.8	1525	4	US-09-540-245A-2	Sequence 2, Appli
6	395.5	11.8	1525	4	US-09-540-153-2	Sequence 2, Appli
7	381	11.3	1523	4	US-09-182-024A-2	Sequence 2, Appli
8	379	11.3	649	4	US-09-188-930-305	Sequence 305, App
9	372	11.1	673	4	US-09-063-950-2	Sequence 2, Appli
10	350	10.4	1480	3	US-09-191-647-7	Sequence 7, Appli
11	350	10.4	1480	4	US-09-540-245A-7	Sequence 7, Appli
12	350	10.4	1480	4	US-09-540-153-7	Sequence 7, Appli
13	350	10.4	1480	5	PCT-US91-09055-2	Sequence 2, Appli
14	347	10.3	1480	4	US-09-182-024A-5	Sequence 5, Appli
15	335	10.0	605	4	US-09-063-950-5	Sequence 5, Appli
16	329.5	9.8	560	3	US-08-592-500-2	Sequence 2, Appli
17	329.5	9.8	560	3	US-08-195-006-2	Sequence 2, Appli
18	329.5	9.8	560	4	US-09-063-950-4	Sequence 4, Appli
19	329.5	9.8	560	5	PCT-US94-07644A-2	Sequence 2, Appli
20	320	9.5	605	1	US-08-190-802A-49	Sequence 49, Appl
21	320	9.5	605	4	US-08-477-346-49	Sequence 49, Appl
22	320	9.5	605	4	US-08-473-089-49	Sequence 49, Appl
23	312	9.3	603	1	US-08-190-802A-50	Sequence 50, Appl
24	312	9.3	603	4	US-08-477-346-50	Sequence 50, Appl
25	312	9.3	603	4	US-08-473-089-50	Sequence 50, Appl
26	303.5	9.0	353	3	US-08-986-485-6	Sequence 6, Appli
27	302	9.0	282	1	US-08-442-063A-45	Sequence 45, Appl
28	302	9.0	307	1	US-08-442-063A-48	Sequence 48, Appl
29	302	9.0	333	1	US-08-442-063A-27	Sequence 27, Appl
30	302	9.0	342	1	US-08-272-919-2	Sequence 2, Appli
31	302	9.0	342	1	US-08-619-916-2	Sequence 2, Appli
32	302	9.0	342	5	PCT-US95-08542-2	Sequence 2, Appli
33	302	9.0	359	1	US-08-303-238-4	Sequence 4, Appli
34	302	9.0	359	4	US-08-458-834-4	Sequence 4, Appli
35	299	8.9	231	3	US-08-986-485-7	Sequence 7, Appli
36	295	8.8	353	6	5340934-4	Patent No. 5340934
37	293	8.7	222	5	PCT-US91-09055-3	Sequence 3, Appli
38	290.5	8.6	610	1	US-07-821-717B-6	Sequence 6, Appli
39	290.5	8.6	610	1	US-08-119-262B-6	Sequence 6, Appli
40	290.5	8.6	610	1	US-08-135-929A-11	Sequence 11, Appl
41	290.5	8.6	610	1	US-08-234-265A-11	Sequence 11, Appl
42	276	8.2	440	3	US-08-985-335-3	Sequence 3, Appli
43	276	8.2	440	4	US-09-410-372-3	Sequence 3, Appli
44	275.5	8.2	368	1	US-08-303-238-3	Sequence 3, Appli
45	275.5	8.2	368	4	US-08-458-834-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-131-648-2
; Sequence 2, Application US/09131648
; Patent No. 6168920
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: EXTRACELLULAR ADHESIVE PROTEINS
; FILE REFERENCE: PF-0576 US
; CURRENT APPLICATION NUMBER: US/09/131,648
; CURRENT FILING DATE: 1998-08-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 2687731
US-09-131-648-2

Query Match 12.8%; Score 432; DB 4; Length 708;
Best Local Similarity 24.0%; Pred. No. 1.6e-28;
Matches 166; Conservative 104; Mismatches 250; Indels 172; Gaps 21;

```
Qy      25 PLLV-VLLALQL-LVVAGLVRAQTCPSVCSCS----------NQFSKVICVRKNLREVP 71
           || : ||| | :  :|  : :   || :|:|            : | | |    |   |
Db       5 PLRIHVLLGLAITTLVQAVDKKVDCPRLCTCEIRPWFTPRSIYMEASTVDCNDLGLLTFP 64

Qy      72 DGISTNTRLLNLHENQIQIIK------VN----------------SFKHLRHLEILQLS 108
             :  ||::| |  | | |  |:      ||                : | :  |  : |
Db      65 ARLPANTQILLLQTNNIAKIEYSTDFPVNLTGLDLSQNNLSSVTNINVKKMPQLLSVYLE 124

Qy     109 RNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAF 168
            | :  :     : |:||  | :  | |:||  |||: |  |  | | :| :: | |  |
Db     125 ENKLTELPEKCLSELSNLQELYINHNLLSTISPGAFIGLHNLLRLHLNSNRLQMINSKWF 184

Qy     169 NRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPN-------------- 214
           : :|:|  | :||   :  | :  |: | ||| | :|  || |||:
Db     185 DALPNLEILMIGE-NPIIRIKDMNFKPLINLRSLVIAGINLTEIPDNALVGLENLESISF 243

Qy     215 ------------LTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWM--------IQS- 253
                       |  :: |  |||: | :: || | |  ::||::| :         | |
Db     244 YDNRLIKVPHVALQKVVNLKFLDLNKNPINRIRRGDFSNMLHLKELGINNMPELISIDSL 303

Qy     254 -----------------QIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERI 296
```

```
                    ::  |  |||  |  |  : |  | |: | |     | :|: |
Db    304 AVDNLPDLRKIEATNNPRLSYIHPNAFFRLPKLESLMLNSNALSALYHGTIESLPNLKEI 363

Qy    297 HLHHNPWNCNCDILWLSW---WIKDMAPSNTACCARCNTPPNLKGRYIGELD-QNYFTCY 352
          :| ||  |:| | |::     |: | | :  |      ||  :|: : ::  ::
Db    364 SIHSNPIRCDCVIRWMNMNKTNIRFMEPDSLFCV----DPPEFQGQNVRQVHFRDMMEIC 419

Qy    353 APVIVEP--PADLNVTEGMAAELKCRASTS-LTSVSWITPNGTVMTHGAYKVRIAVLSDG 409
           |:|     |:|||  |      |||:      : ||||:|  :        :  | |:|
Db    420 LPLIAPESFPSNLNVEAGSYVSFHCRATAEPQPEIYWITPSGQKLLPNTLTDKFYVHSEG 479

Qy    410 TLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEA 469
          ||:   || :: |:|||: :| ||    |  : |                      | :
Db    480 TLDINGVTPKEGGLYTCIATNLVGADLKSVMIKV----------------------DGS 516

Qy    470 RTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKII 529
            |||                          :  | : || :     : |: : |
Db    517 FPQDNN-------------------------GSLNIKIRDIQA-----NSVLVSWKA- 543

Qy    530 IGCFVAITLMAAVMLVIFYKMRKQHHRQNHHAPTRTVEIINVDDEITGDTPMESHLPMPA 589
                :  | ::|    | |    |  |:   |:  |:: |:   :   |  :  : :|
Db    544 ----SSKILKSSVKWTAFVKTENSHAAQSARIPS-DVKVYNL-THLNPSTEYKICIDIPT 597

Qy    590 IEHEHLN----------HYNSYKSPFNHTTTV 611
            |  ::          | :  :   |:|||:
Db    598 IYQKNRKKCVNVTTKGLHPDQKEYEKNNTTTL 629


RESULT   2
US-08-986-485-5
; Sequence 5, Application US/08986485
; Patent No. 6046030
;  GENERAL INFORMATION:
;    APPLICANT:  WU, SHUJIAN
;    APPLICANT:  SWEET, RAYMOND
;    APPLICANT:  TRUNEH, ALEMSEGED
;    TITLE OF INVENTION:  A HUMAN LIG-1 HOMOLOG (HLIG-1)
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RATNER & PRESTIA
;      STREET:  P.O. BOX 980
;      CITY:  VALLEY FORGE
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/986,485
;      FILING DATE:  08-DEC-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/059,448
;      FILING DATE:  22-SEP-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PRESTIA, PAUL F
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  GH-70264
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1091 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-986-485-5
```

```
  Query Match             12.4%;  Score 417.5;  DB 3;  Length 1091;
  Best Local Similarity   23.1%;  Pred. No. 5.3e-27;
  Matches  156;  Conservative   87;  Mismatches  208;  Indels  223;  Gaps   19;

Qy      81 LNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIP 140
           |:|: |:|::|:  :|: |  ||:|:| ||:|  :  ||| ||: :: | |  | |  :
Db    218 LDLNRNRIRLIEGLTFQGLDSLEVLRLQRNNISRLTDGAFWGLSKMHVLHLEYNSLVEVN 277

Qy    141 NGAFVYLSKLKELWLRNNPIESIPS------------YAFNRIPSLRRLDLGELKRL-- 185
           :|:   |: | :| | || |  |                :|| :  |    | ||  |
Db    278 SGSLYGLTALHQLHLSNNSISRIQRDGWSFCQKLHELILSFNNLTRLDEESLAELSSLSI 337

Qy    186 --------SYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPG 237
                  |:|:|||||:|| :|| |:              ||  ::||         |
Db    338 LRLSHNAISHIAEGAFKGLKSLRVLD----------------LDHNEISGTIEDT--SG 378

Qy    238 SFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIH 297
           :| || :| || :  ::|: : : ||  |:||  :||  | :  :  | |  : :|: ::
Db    379 AFTGLDNLSKLTLFGNKIKSVAKRAFSGLESLEHLNLGENAIRSVQFDAFAKMKNLKELY 438

Qy    298 LHHNPWNCNCDILWLSWWI--KDMAPSNTACCARCNTPPNLKGRYI-GELDQNY------ 348
           :    : |:| : ||  |:  : :    || ||    | :|||: |   |  ::
Db    439 ISSESFLCDCQLKWLPPWLMGRMLQAFVTATCAH---PESLKGQSIFSVLPDSFVCDDFP 495

Qy    349 --------------------FTCYA---------------------------------- 353
                              ||| |
Db    496 KPQIITQPETTMAVVGKDIRFTCSAASSSSSPMTFAWKKDNEVLANADMENFAHVRAQDG 555

Qy    354 ------------------------------------------PVIVEPPADLNVTEG 368
                                                     |   : | |: :  |
Db    556 EVMEYTTILHLRHVTFGHEGRYQCIITNHFGSTYSHKARLTVNVLPSFTKIPHDIAIRTG 615

Qy    369 MAAELKCRASTSLT-SVSWITPNGTVMTHGAYKVRIAVLSDGTLNF-TNVTVQDTGMYTC 426
           | |:| |:       ::|    ||      | : |: |: |  :  | |:| : | |:|:|
Db    616 TTARLECAATGHPNPQIAWQKDGGTDFP-AARERRMHVMPDDDVFFITDVKIDDMGVYSC 674

Qy    427 MVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNV--------GP 478
             || |: :|:||| |                |     | :|   |    |         |
Db    675 TAQNSAGSVSANATLTVLE-----------TPSLAVPLEDRVVTVGETVAFQCKATGSP 722

Qy    479 TPVVDWETTNVTTSLT------------------------------PQSTRSTEKTFTI 507
           || : |       |||                              |  |      :|
Db    723 TPRITWLKGGRPLSLTERHHFTPGNQLLVVQNVMIDDAGRYTCEMSNPLGTERAHSQLSI 782

Qy    508 PVTDINSGIPGIDEVMKTTKII-IGCFVAITLMAAVMLVIFYKMRKQHHRQNHHAPTRTV 566
             |      ||  :   |  |  |    :| | : | : | |: ||   :   :: | |
Db    783 LPT------PGCRKDGTTVGIFTIAVVCSIVLTSLVWVCIIYQTRK---KSEEYSVTNTD 833

Qy    567 EIINVDDEITGDTP 580
           | |     :  | |
Db    834 ETI-----VPPDVP 842


RESULT   3
US-08-986-485-2
; Sequence 2, Application US/08986485
; Patent No. 6046030
;  GENERAL INFORMATION:
;    APPLICANT:  WU, SHUJIAN
;    APPLICANT:  SWEET, RAYMOND
;    APPLICANT:  TRUNEH, ALEMSEGED
;    TITLE OF INVENTION:  A HUMAN LIG-1 HOMOLOG (HLIG-1)
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RATNER & PRESTIA
;      STREET:  P.O. BOX 980
;      CITY:  VALLEY FORGE
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/986,485
;      FILING DATE:  08-DEC-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/059,448
;      FILING DATE:  22-SEP-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PRESTIA, PAUL F
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  GH-70264
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1101 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-986-485-2


  Query Match             11.9%;  Score 401;  DB 3;  Length 1101;
  Best Local Similarity   22.0%;  Pred. No. 1.4e-25;
  Matches  172;  Conservative  111;  Mismatches  265;  Indels  234;  Gaps   27;

Qy      26 LLVVLLALQLLVVAGLVRAQTCPSVCSCSNQ-----FSKVICVRKNLREVPDGISTNTRL 80
           | ::|| |: :  |   ||  | : |:|:         : |  : |  :|  : : ||
Db      21 LWLLLLRLEPVTAAAGPRA-PCAAACTCAGDPCTCAGDSLDCGGRGLAALPGDLPSWTRS 79

Qy      81 LNL--------------------------HE----------------------------- 85
           |||                          ||
Db      80 LNLSYNKLAEIDPAGFEDLPNLQEVYLNNHELTAVASLGAGSSQVVALFLQQQQNRSLDG 139

Qy      86 ----------------NQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLA-NLNT 128
                           | |  ::   | |   :: | |: | | |:|:|||:||: :| |
Db     140 SQLKAYLSLEVLDLNLNNITEVRNTYFPHGPPIKELNLAGNRIGTLELGAFDGLSRSLLT 199

Qy     129 LELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYI 188
           | |  ||:| :|  ||  | :| :| |  | |  |    |  : ||  | | :   :| :
Db     200 LRLSKNRITQLPVRAF-KLPRLTQLDLNRNRIRLIEGLTFQGLNSLEVLKL-QRNNISKL 257

Qy     189 SEGAFEGLSNLRYLNLAMCNLREI---------------------------------- 212
           ::||| ||| :  |:|   :| |:
Db     258 TDGAFWGLSKMHVLHLEYDSLVEVNSGSLYGLTALHQLHLSNNSIARIHRKGWSFCQKLH 317

Qy     213 -------------PNLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQ------------ 246
                        :|  |  |  | || | :| |  |:|:||  |:
Db     318 ELVLSFNNLTRLDEESLAELSSLSVLRLSHNSISHIAEGAFKGLRSLRVLDLDHNEISGT 377

Qy     247 ----------------KLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPL 290
                           || :  ::|: : : ||  |: |  :||  | :  :  | |  :
Db     378 IEDTSGAFSGLEFGHSKLTLFGNKIKSVAKRAFSGLEGLEHLNLGGNAIRSVQFDAFVKM 437

Qy     291 HHLERIHLHHNPWNCNCDILWLSWWI--KDMAPSNTACCARCNTPPNLKGRYIGELDQNY 348
           :|: :|:  : : |:| : ||  |:  : :    || ||    | :|||: |  :
Db     438 KNLKELHISSDSFLCDCQLKWLPPWLIGRMLQAFVTATCAH---PESLKGQSIFSVPPES 494

Qy     349 FTC---YAPVIVEPPADLNVTEGMAAELKCRASTSLTS---VSWITPNGTVMTHGAYKVR 402
           | |     | |:  |       |  | |::| :|    :|   |  |:|:   :
Db     495 FVCDDFLKPQIITQPETTMAMVGKDIRFTCSAASSSSSPMTFAWKKDN-EVLTNADMENF 553

Qy     403 IAV-LSDG-------TLNFTNVTVQDTGMYTCMVSNSVGNT---TASATLNVTAA-TTTP 450
           : |   ||         |:   ||    | | |:::|  |:|    |  |:||  : | ||
Db     554 VHVHAQDGEVMEYTTILHLRQVTFGHEGRYQCVITNHFGSTYSHKARLTVNVLPSFTKTP 613

Qy     451 FSYFSTVTVETMEPSQDEARTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIP-- 508
                 :|: |   :: |   | :   | | : |:    |      :  |       :|
Db     614 HD----ITIRTTTVARLECAATGH---PNPQIAWQKDGGTDFPAARERR----MHVMPDD 662

Qy     509 ----VTDINSGIPGIDE-VMKTTKIIIGCFVAITLMAAVMLVIFYKMRKQHHRQNHHAPT 563
               :||:     |:     : :   |     :|::    ||:  : |
Db     663 DVFFITDVKIDDAGVYSCTAQNSAGSISANATLTVLETPSLVVPLEDR------------ 710

Qy     564 RTVEIINVDDEI------TGDTP-----MESHLPMPAIEHEHLNHYNSYKSPFNHTTTVN 612
               :::| : :      ||: |      :   |:   |  ||      :| |   |
Db     711 ----VVSVGETVALQCKATGNPPPRITWFKGDRPLSLTERHHL-------TPDNQLLVVQ 759

Qy     613 TI 614
             :
Db     760 NV 761


RESULT   4
US-09-191-647-2
; Sequence 2, Application US/09191647
; Patent No. 6046015
; GENERAL INFORMATION:
;  APPLICANT: Goodman, Corey
;  APPLICANT: Kid, Thomas
;  APPLICANT: Brose, Katja
;  APPLICANT: Tessier-Lavigne, Marc
;  TITLE OF INVENTION: Modulating Robo: Ligand Interactions
;  FILE REFERENCE: B98-031-3
;  CURRENT APPLICATION NUMBER: US/09/191,647
;  CURRENT FILING DATE: 1998-11-13
;  EARLIER APPLICATION NUMBER: 60/065,544
;  EARLIER FILING DATE: 1997-11-14
;  EARLIER APPLICATION NUMBER: 60/081,057
;  EARLIER FILING DATE: 1998-04-07
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1525
;   TYPE: PRT
;   ORGANISM: human
US-09-191-647-2


  Query Match             11.8%;  Score 395.5;  DB 3;  Length 1525;
  Best Local Similarity   25.1%;  Pred. No. 6.7e-25;
  Matches  129;  Conservative   51;  Mismatches  171;  Indels  163;  Gaps    9;

Qy      30 LLALQL---LVVAGLVRAQTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHEN 86
           :|:| |   | :   |  | ||: ||||   | | |    || ||  |  ||  |:|: |
Db       8 MLSLSLGLVLAILNKVAPQACPAQCSCSG--STVDCHGLALRSVPRNIPRNTERLDLNGN 65

Qy      87 QIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVY 146
           |  |    |  |||| :|||  | | ||| |||  |  |  | |  | |    |   |:
Db      66 NITRITKTDFAGLRHLRVLQLMENKISTIERGAFQDLKELERLRLNRNHLQLFPELLFLG 125

Qy     147 LSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNL-- 204
           :||  | | |  | |::||  ||      :: | | :  ::| | :|||  | :|  | |
Db     126 TAKLYRLDLSENQIQAIPRKAFRGAVDIKNLQL-DYNQISCIEDGAFRALRDLEVLTLNN 184

Qy     205 ------------------------------------------------------------ 204

Db     185 NNITRLSVASFNHMPKLRTFRLHSNNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPSHLRG 244

Qy     205 ---------------------------------AMC------------NLREIPN---- 214
                                               | |            | |||
Db     245 HNVAEVQKREFVCSDEEEGHQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPTNLPE 304

Qy     215 -------------------LTPLIKLDELDLSGNHLSAIRPGSFQ--------------- 240
                              :|  ||  :||| | :| : | :||
Db     305 TITEIRLEQNTIKVIPPGAFSPYKKLRRIDLSNNQISELAPDAFQGLRSLNSLVLYGNKI 364

Qy     241 ---------GLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLH 291
                    ||  || | :  ::|  :  :|| :| :|  ::|  | |  :    |:||
```

```
Db    365 TELPKSLFEGLFSLQLLLLNANKINCLRVDAFQDLHNLNLLSLYDNKLQTIAKGTFSPLR 424

Qy    292 HLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTC 351
          :: :||  ||: |:| : ||: ::    |  |:  ||| :|  |  : ||::    | |
Db    425 AIQTMHLAQNPFICDCHLKWLADYL-HTNPIETS-GARCTSPRRLANKRIGQIKSKKFRC 482

Qy    352 YAPVIVEPPADLNVTEGMAAELKCRASTSLTSVS 385
                      :    :|   |||   :    |
Db    483 SGTEDYRSKLSGDCFADLACPEKCRCEGTTVDCS 516


RESULT   5
US-09-540-245A-2
; Sequence 2, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
;  APPLICANT:  Goodman, Corey
;  APPLICANT:  Kid, Thomas
;  APPLICANT:  Brose, Katja
;  APPLICANT:  Tessier-Lavigne, Marc
;  TITLE OF INVENTION:  Modulating Robo: Ligand Interactions
;  FILE REFERENCE:  B98-031-3
;  CURRENT APPLICATION NUMBER: US/09/540,245A
;  CURRENT FILING DATE:  2000-03-31
;  PRIOR APPLICATION NUMBER:  60/065,544
;  PRIOR FILING DATE:  1997-11-14
;  PRIOR APPLICATION NUMBER:  60/081,057
;  PRIOR FILING DATE:  1998-04-07
;  NUMBER OF SEQ ID NOS:  20
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1525
;   TYPE: PRT
;   ORGANISM: human
US-09-540-245A-2


  Query Match           11.8%;  Score 395.5;  DB 4;  Length 1525;
  Best Local Similarity   25.1%;  Pred. No. 6.7e-25;
  Matches  129;  Conservative   51;  Mismatches  171;  Indels  163;  Gaps    9;

Qy     30 LLALQL---LVVAGLVRAQTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHEN 86
          :|:| |   | :   |  | ||: ||||   | | |    || ||  |  ||  |:|: |
Db      8 MLSLSLGLVLAILNKVAPQACPAQCSCSG--STVDCHGLALRSVPRNIPRNTERLDLNGN 65

Qy     87 QIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVY 146
           |  |    |  |||| :|||  | | ||| |||  |  |  | |  |  |    |:
Db     66 NITRITKTDFAGLRHLRVLQLMENKISTIERGAFQDLKELERLRLNRNHLQLFPELLFLG 125

Qy    147 LSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNL-- 204
          :||  | |  | |::||  ||     :: | | :  ::| | :|||  | :|  | |
Db    126 TAKLYRLDLSENQIQAIPRKAFRGAVDIKNLQL-DYNQISCIEDGAFRALRDLEVLTLNN 184

Qy    205 ------------------------------------------------------------ 204

Db    185 NNITRLSVASFNHMPKLRTFRLHSNNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPSHLRG 244

Qy    205 ---------------------------------AMC-----------NLREIPN---- 214
                                              | |           | |||
Db    245 HNVAEVQKREFVCSDEEEGHQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPTNLPE 304

Qy    215 ------------------LTPLIKLDELDLSGNHLSAIRPGSFQ--------------- 240
                            :|  ||  :||| | :| : | :||
Db    305 TITEIRLEQNTIKVIPPGAFSPYKKLRRIDLSNNQISELAPDAFQGLRSLNSLVLYGNKI 364

Qy    241 ---------GLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLH 291
                  ||  || | :  ::|  :  :|| :| :|  ::|  | |  :    |:||
Db    365 TELPKSLFEGLFSLQLLLLNANKINCLRVDAFQDLHNLNLLSLYDNKLQTIAKGTFSPLR 424

Qy    292 HLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTC 351
          :: :||  ||: |:| : ||: ::    |  |:  ||| :|  |  : ||::    | |
Db    425 AIQTMHLAQNPFICDCHLKWLADYL-HTNPIETS-GARCTSPRRLANKRIGQIKSKKFRC 482

Qy    352 YAPVIVEPPADLNVTEGMAAELKCRASTSLTSVS 385
                      :    :|   |||   :    |
Db    483 SGTEDYRSKLSGDCFADLACPEKCRCEGTTVDCS 516


RESULT   6
US-09-540-153-2
; Sequence 2, Application US/09540153
; Patent No. 6270995
; GENERAL INFORMATION:
;  APPLICANT:  Goodman, Corey
;  APPLICANT:  Kid, Thomas
;  APPLICANT:  Brose, Katja
;  APPLICANT:  Tessier-Lavigne, Marc
;  TITLE OF INVENTION:  Modulating Robo: Ligand Interactions
;  FILE REFERENCE:  B98-031-3
;  CURRENT APPLICATION NUMBER: US/09/540,153
;  CURRENT FILING DATE:  2000-03-31
;  PRIOR APPLICATION NUMBER:  09/191,647
;  PRIOR FILING DATE:  1998-11-13
;  PRIOR APPLICATION NUMBER:  60/081,057
;  PRIOR FILING DATE:  1998-04-07
;  NUMBER OF SEQ ID NOS:  14
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1525
;   TYPE: PRT
;   ORGANISM: human
US-09-540-153-2


  Query Match           11.8%;  Score 395.5;  DB 4;  Length 1525;
  Best Local Similarity   25.1%;  Pred. No. 6.7e-25;
  Matches  129;  Conservative   51;  Mismatches  171;  Indels  163;  Gaps    9;

Qy     30 LLALQL---LVVAGLVRAQTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHEN 86
          :|:| |   | :   |  | ||: ||||   | | |    || ||  |  ||  |:|: |
Db      8 MLSLSLGLVLAILNKVAPQACPAQCSCSG--STVDCHGLALRSVPRNIPRNTERLDLNGN 65

Qy     87 QIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVY 146
           |  |    |  |||| :|||  | | ||| |||  |  |  | |  |  |    |:
Db     66 NITRITKTDFAGLRHLRVLQLMENKISTIERGAFQDLKELERLRLNRNHLQLFPELLFLG 125

Qy    147 LSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNL-- 204
          :||  | |  | |::||  ||     :: | | :  ::| | :|||  | :|  | |
Db    126 TAKLYRLDLSENQIQAIPRKAFRGAVDIKNLQL-DYNQISCIEDGAFRALRDLEVLTLNN 184

Qy    205 ------------------------------------------------------------ 204

Db    185 NNITRLSVASFNHMPKLRTFRLHSNNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPSHLRG 244

Qy    205 ---------------------------------AMC-----------NLREIPN---- 214
                                              | |           | |||
Db    245 HNVAEVQKREFVCSDEEEGHQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPTNLPE 304

Qy    215 ------------------LTPLIKLDELDLSGNHLSAIRPGSFQ--------------- 240
                            :|  ||  :||| | :| : | :||
Db    305 TITEIRLEQNTIKVIPPGAFSPYKKLRRIDLSNNQISELAPDAFQGLRSLNSLVLYGNKI 364

Qy    241 ---------GLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLH 291
                  ||  || | :  ::|  :  :|| :| :|  ::|  | |  :    |:||
Db    365 TELPKSLFEGLFSLQLLLLNANKINCLRVDAFQDLHNLNLLSLYDNKLQTIAKGTFSPLR 424

Qy    292 HLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTC 351
          :: :||  ||: |:| : ||: ::    |  |:  ||| :|  |  : ||::    | |
Db    425 AIQTMHLAQNPFICDCHLKWLADYL-HTNPIETS-GARCTSPRRLANKRIGQIKSKKFRC 482

Qy    352 YAPVIVEPPADLNVTEGMAAELKCRASTSLTSVS 385
                      :    :|   |||   :    |
Db    483 SGTEDYRSKLSGDCFADLACPEKCRCEGTTVDCS 516


RESULT   7
```

```
US-09-182-024A-2
; Sequence 2, Application US/09182024A
; Patent No. 6342370
; GENERAL INFORMATION:
;  APPLICANT:  Connolly, Timothy
;  APPLICANT:  Rajput, Bhanu
;  TITLE OF INVENTION:  Human Slit Polypeptide and Polynucleotides Encoding
;  TITLE OF INVENTION:  Same
;  FILE REFERENCE:  640100-271
;  CURRENT APPLICATION NUMBER: US/09/182,024A
;  CURRENT FILING DATE:  1998-10-29
;  PRIOR APPLICATION NUMBER:  60/063,946
;  PRIOR FILING DATE:  1997-10-31
;  PRIOR APPLICATION NUMBER:  60/096,420
;  PRIOR FILING DATE:  1998-08-13
;  NUMBER OF SEQ ID NOS:  5
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 1523
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-182-024A-2


  Query Match           11.3%;  Score 381;  DB 4;  Length 1523;
  Best Local Similarity  25.0%;  Pred. No. 1.2e-23;
  Matches  138;  Conservative   56;  Mismatches  183;  Indels  176;  Gaps   13;

Qy      31 LALQLLVVAGLVRAQTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQI 90
           |||  |   |      |  ||: |:||   : | |    || || ||  |   |:|  | |
Db      18 LALALASVLSGPPAVACPTKCTCS--AASVDCHGLGLRAVPRGIPRNAERLDLDRNNITR 75

Qy      91 IKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKL 150
           |    |  |::| :| |  | :  || |||  |  |  | |  |:|  :|   |    ||
Db      76 ITKMDFAGLKNLRVLHLEDNQVSVIERGAFQDLKQLERLRLNKNKLQVLPELLFQSTPKL 135

Qy     151 KELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNL------ 204
             | |  | |: ||  ||  |  :: | | :   :| | :|||  | :|  | |
Db     136 TRLDLSENQIQGIPRKAFRGITDVKNLQL-DNNHISCIEDGAFRALRDLEILTLNNNNIS 194

Qy     205 ------------------------------------------------------------ 204

Db     195 RILVTSFNHMPKIRTLRLHSNHLYCDCHLAWLSDWLRQRRTVGQFTLCMAPVHLRGFNVA 254

Qy     205 ----------------AMCN-----------------------LREIP-NL--------- 215
                           ||                        | ||| ||
Db     255 DVQKKEYVCPAPHSEPPSCNANSISCPSPCTCSNNIVDCRGKGLMEIPANLPEGIVEIRL 314

Qy     216 -------------TPLIKLDELDLSGNHLSAIRPGSFQ---------------------- 240
                       |   ||  :|:| | :| | | :||
Db     315 EQNSIKAIPAGAFTQYKKLKRIDISKNQISDIAPDAFQGLKSLTSLVLYGNKITEIAKGL 374

Qy     241 --GLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 298
             ||: || | :  ::|  :  | | :||:|  ::|  | |  :   || ||  :: :||
Db     375 FDGLVSLQLLLLNANKINCLRVNTFQDLQNLNLLSLYDNKLQTISKGLFAPLQSIQTLHL 434

Qy     299 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVE 358
             ||: |:| : ||: :::|  |  |:  |||::|  |  : | ::    | |
Db     435 AQNPFVCDCHLKWLADYLQD-NPIETS-GARCSSPRRLANKRISQIKSKKFRCSGSEDYR 492

Qy     359 PPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYK-VRI-----AVLSDGTLN 412
                 :  |||              ||::     | |||        ::|  ||
Db     493 SRFSSECFMDLVCPEKCRC------------EGTIVDCSNQKLVRIPSHLPEYVTDLRLN 540

Qy     413 FTNVTV-QDTGMY 424
             |:| : ||::
Db     541 DNEVSVLEATGIF 553


RESULT   8
US-09-188-930-305
; Sequence 305, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT: Strachan, Lorna
;  APPLICANT: Sleeman, Matthew
;  APPLICANT: Onrust, Rene
;  APPLICANT: Murison, James Greg
;  TITLE OF INVENTION: Compositions Isolated From Skin Cells
;  TITLE OF INVENTION: and Methods For Their Use
;  FILE REFERENCE: 11000.1011c1
;  CURRENT APPLICATION NUMBER: US/09/188,930A
;  CURRENT FILING DATE: 1998-11-09
;  NUMBER OF SEQ ID NOS: 348
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 305
;   LENGTH: 649
;   TYPE: PRT
;   ORGANISM: Mouse
US-09-188-930-305


  Query Match           11.3%;  Score 379;  DB 4;  Length 649;
  Best Local Similarity  23.4%;  Pred. No. 4.7e-24;
  Matches  165;  Conservative  109;  Mismatches  252;  Indels  178;  Gaps   26;

Qy      13 MIGPRFNRALFDPLLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKVICVRKNLREVPD 72
           || | :  :||  |:   : |   |    | |::||||| |   |  : |  ::|  :|
Db       1 MISPAW--SLF--LIGTKIGLFFQVAPLSVVAKSCPSVCRCDAGF--IYCNDRSLTSIPV 54

Qy      73 GISTNTRLLNLHENQIQIIKVNS-FKHL--------------------RHLEILQLSRN 110
           ||  :   | |  |||  : : |  |:|                    :::: | | |
Db      55 GIPEDATTLYLQNNQINNVGIPSDLKNLLKVQRIYLYHNSLDEFPTNLPKYVKELHLQEN 114

Qy     111 HIRTIEIGAFNGLANLNTLELFDNRLT--TIPNGAFVYLSKLKELWLRNNPIESIPSYAF 168
           :||||   : : :  |  | | || ::  :|  |||   :  |: |:|  | : :||
Db     115 NIRTITYDSLSKIPYLEELHLDDNSVSAVSIEEGAFRDSNYLRLLFLSRNHLSTIPGGLP 174

Qy     169 NRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTP-----LIKLDE 223
             |  ||  |     |:| ||  :  ||::|: | |   ||      |       |: | |
Db     175 RTIEELRLDD----NRISTISSPSLHGLTSLKRLVLD-GNLLNNHGLGDKVFFNLVNLTE 229

Qy     224 LDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLP 283
           | |  | |:|  | :  |   |:||::  : | :  |||  |: |  :::::|||: ||
Db     230 LSLVRNSLTA-APVNLPG-TSLRKLYLQDNHINRVPPNAFSYLRQLYRLDMSNNNLSNLP 287

Qy     284 HDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGE 343
             :|  | :: :: | :|||  | | : |:  |:: :           |  |  ::|  | :
Db     288 QGIFDDLDNITQLILRNNPWYCGCKMKWVRDWLQSLPVKVNVRGLMCQAPEKVRGMAIKD 347

Qy     344 LDQNYFTCY-------------------------APVIVEPP-------ADLNVTEGMA 370
           |   | |                          |||  :|        |  |   :
Db     348 LSAELFDCKDSGIVSTIQITTAIPNTAYPAQGQWPAPVTKQPDIKNPKLIKDQRTTGSPS 407

Qy     371 AELKCRASTSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSN 430
             :      | | :|  :||:     | :::: : :               | :    :
Db     408 RK------TILITVKSVTPD---TIHISWRLALPM--------------TALRLSWL-- 441

Qy     431 SVGNTTASATLNVTAATTTPFSYFST---------VTVETMEPSQ----DEARTTDNNVG 477
             :|:: |  ::  |  |      | :  || |     ||
Db     442 KLGHSPAFGSITETIVTGERSEYLVTALEPESPYRVCMVPMETSNLYLFDE--------- 492

Qy     478 PTPV-VDWETT-----NVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTKIIIG 531
           ||| :: :|      | ||:|  :  :   |   :|:                  |||
Db     493 -TPVCIETQTAPLRMYNPTTTLNREQEKEPYKNPNLPLA----------------AIIG 534

Qy     532 CFVAITLMAAVMLVIFYKMR------------KQHHRQNHHAPTRT-------------V 566
             ||:  :| : || :|  |             |   |:: :|   |
Db     535 GAVALVSIALLALVCWYVHRNGSLFSRNCAYSKGRRRKDDYAEAGTKKDNSILEIRETSF 594

Qy     567 EIINVDDEITGDTPMESHLPMPAIEHEHLNHYNSYKSPFNHTTT 610
           ::: : :|         |   |       |  | ||:  : :::
Db     595 QMLPISNEPISKEEFVIHTIFPP------NGMNLYKNNLSESSS 632
```

```
RESULT   9
US-09-063-950-2
; Sequence 2, Application US/09063950C
; Patent No. 6225085
; GENERAL INFORMATION:
;  APPLICANT: Holtzman, Douglas A.
;  TITLE OF INVENTION: NOVEL LRSG PROTEIN AND NUCLEIC ACID MOLECULES AND USES
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: MEI-019
;  CURRENT APPLICATION NUMBER: US/09/063,950C
;  CURRENT FILING DATE: 1998-04-21
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 673
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-063-950-2


  Query Match             11.1%;  Score 372;  DB 4;  Length 673;
  Best Local Similarity   30.4%;  Pred. No. 2e-23;
  Matches  123;  Conservative   58;  Mismatches  156;  Indels   68;  Gaps   13;

Qy      25 PLLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLH 84
           |||: ||   ||: |    | ||| | || |   | |  :    ||  :  :|  | :
Db       6 PLLLPLL---LLLALG-PGVQGCPSGCQCS-QPQTVFCTARQGTTVPRDVPPDTVGLYVF 60

Qy      85 ENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAF 144
           || | ::   ||  |  |:::| ||:| | ::  | |  ||||: |:|  |||  | |  |
Db      61 ENGITMLDAGSFAGLPGLQLLDLSQNQIASLPSGVFQPLANLSNLDLTANRLHEITNETF 120

Qy     145 VYLSKLKELWLRNNPIESIPSYAFN--------------------RIPSLRRLDLGELK 183
             | :|: |:|  | |  |   ||:                    |:| |  |||
Db     121 RGLRRLERLYLGKNRIRHIQPGAFDTLDRLLELKLQDNELRALPPLRLPRLLLLDLSHNS 180

Qy     184 RLSYISEGAFEGLSNLRYLNLAMCNLREIPN--LTPLIKLDELDLSGNHLSAIRPGSFQG 241
           |:   |      :|:  | ||   |:::     : |  | :||:| | |  : |   :|
Db     181 LLAL--EPGILDTANVEALRLAGLGLQQLDEGLFSRLRNLHDLDVSDNQLERV-PPVIRG 237

Qy     242 LMHLQKLWMI-QSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHH 300
           |  | :| :   ::|  :       | :| |:::: :|  || ||      |  :
Db     238 LRGLTRLRLAGNTRIAQLRPEDLAGLAALQELDVSNLSLQALPGDLSGLFPRLRLLAAAR 297

Qy     301 NPWNCNCDILWLSWWIKD-----MAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYA-- 353
           ||:|| | : |   |:::      :|  |    ||: ||   || : |||   | | |
Db     298 NPFNCVCPLSWFGPWVRESHVTLASPEET----RCHFPPKNAGRLLLELDYADFGCPATT 353

Qy     354 ---------PVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITP 389
                   ||: || |           |:|| : :|::|
Db     354 TTATVPTTRPVVREPTA--------------LSSSL-APTWLSP 382


RESULT  10
US-09-191-647-7
; Sequence 7, Application US/09191647
; Patent No. 6046015
; GENERAL INFORMATION:
;  APPLICANT: Goodman, Corey
;  APPLICANT: Kid, Thomas
;  APPLICANT: Brose, Katja
;  APPLICANT: Tessier-Lavigne, Marc
;  TITLE OF INVENTION: Modulating Robo: Ligand Interactions
;  FILE REFERENCE: B98-031-3
;  CURRENT APPLICATION NUMBER: US/09/191,647
;  CURRENT FILING DATE: 1998-11-13
;  EARLIER APPLICATION NUMBER: 60/065,544
;  EARLIER FILING DATE: 1997-11-14
;  EARLIER APPLICATION NUMBER: 60/081,057
;  EARLIER FILING DATE: 1998-04-07
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 1480
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-191-647-7


  Query Match             10.4%;  Score 350;  DB 3;  Length 1480;
  Best Local Similarity   23.9%;  Pred. No. 5.1e-21;
  Matches  109;  Conservative   65;  Mismatches  152;  Indels  130;  Gaps    9;

Qy      46 TCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEIL 105
           :||  | |::    | |  |:|  ||  :  :|  : | :| |  :   ||   | |  :
Db     294 SCPHPCRCADGI--VDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKSFSSFRRLRRI 351

Qy     106 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 165
            || |:|  |   | :||  | || |: |::  :|:| |  |  |: | |  | |  |
Db     352 DLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLLNANEISCIRK 411

Qy     166 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM------CNLR--------- 210
           ||  : ||  | | :   :  :: | |: : ::: ::||       ||||
Db     412 DAFRDLHSLSLLSLYD-NNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRWLADYLHKN 470

Qy     211 ------------------------------------------------------------ 210

Db     471 PIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKLSGECRMDSDCPAMCHCEGTTV 530

Qy     211 --------EIPNLTPL--------------IKLD----------ELDLSGNHLSAIRPGS 238
                   |||   ||              | |           :|:|  | |: | | :
Db     531 DCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKRNQLTGIEPNA 590

Qy     239 FQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 298
           |:|  |:|:| : :::|: |    |  |  |  :||  | :: :    |  |: |  ::|
Db     591 FEGASHIQELQLGENKIKEISNKMFLGLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNL 650

Qy     299 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVE 358
             ||:|||| : | :  ::  : :  |  |||  |  ::   | :|  : | | :
Db     651 ASNPFNCNCHLAWFAECVRKKSLNGGA--ARCGAPSKVRDVQIKDLPHSEFKCSS----- 703

Qy     359 PPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVM 394
                 :||   :  |  | : |         |||:
Db     704 -----ENSEGCLGDGYCPPSCTCT--------GTVV 726


RESULT  11
US-09-540-245A-7
; Sequence 7, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
;  APPLICANT:  Goodman, Corey
;  APPLICANT:  Kid, Thomas
;  APPLICANT:  Brose, Katja
;  APPLICANT:  Tessier-Lavigne, Marc
;  TITLE OF INVENTION:  Modulating Robo: Ligand Interactions
;  FILE REFERENCE:  B98-031-3
;  CURRENT APPLICATION NUMBER: US/09/540,245A
;  CURRENT FILING DATE:  2000-03-31
;  PRIOR APPLICATION NUMBER:  60/065,544
;  PRIOR FILING DATE:  1997-11-14
;  PRIOR APPLICATION NUMBER:  60/081,057
;  PRIOR FILING DATE:  1998-04-07
;  NUMBER OF SEQ ID NOS:  20
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 1480
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-540-245A-7


  Query Match             10.4%;  Score 350;  DB 4;  Length 1480;
  Best Local Similarity   23.9%;  Pred. No. 5.1e-21;
  Matches  109;  Conservative   65;  Mismatches  152;  Indels  130;  Gaps    9;
```

```
Qy      46 TCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEIL 105
Db     294 SCPHPCRCADGI--VDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKSFSSFRRLRRI 351

Qy     106 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 165
Db     352 DLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLLNANEISCIRK 411

Qy     166 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM------CNLR--------- 210
Db     412 DAFRDLHSLSLLSLYD-NNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRWLADYLHKN 470

Qy     211 ------------------------------------------------------------ 210

Db     471 PIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKLSGECRMDSDCPAMCHCEGTTV 530

Qy     211 --------EIPNLTPL--------------IKLD----------ELDLSGNHLSAIRPGS 238
Db     531 DCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKRNQLTGIEPNA 590

Qy     239 FQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 298
Db     591 FEGASHIQELQLGENKIKEISNKMFLGLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNL 650

Qy     299 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVE 358
Db     651 ASNPFNCNCHLAWFAECVRKKSLNGGA--ARCGAPSKVRDVQIKDLPHSEFKCSS----- 703

Qy     359 PPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVM 394
Db     704 -----ENSEGCLGDGYCPPSCTCT--------GTVV 726
```

RESULT 12
US-09-540-153-7
; Sequence 7, Application US/09540153
; Patent No. 6270995
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,153
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/191,647
; PRIOR FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1480
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-540-153-7

Query Match 10.4%; Score 350; DB 4; Length 1480;
Best Local Similarity 23.9%; Pred. No. 5.1e-21;
Matches 109; Conservative 65; Mismatches 152; Indels 130; Gaps 9;

```
Qy      46 TCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEIL 105
Db     294 SCPHPCRCADGI--VDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKSFSSFRRLRRI 351

Qy     106 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 165
Db     352 DLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLLNANEISCIRK 411

Qy     166 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM------CNLR--------- 210
Db     412 DAFRDLHSLSLLSLYD-NNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRWLADYLHKN 470

Qy     211 ------------------------------------------------------------ 210

Db     471 PIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKLSGECRMDSDCPAMCHCEGTTV 530

Qy     211 --------EIPNLTPL--------------IKLD----------ELDLSGNHLSAIRPGS 238
Db     531 DCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKRNQLTGIEPNA 590

Qy     239 FQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 298
Db     591 FEGASHIQELQLGENKIKEISNKMFLGLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNL 650

Qy     299 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVE 358
Db     651 ASNPFNCNCHLAWFAECVRKKSLNGGA--ARCGAPSKVRDVQIKDLPHSEFKCSS----- 703

Qy     359 PPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVM 394
Db     704 -----ENSEGCLGDGYCPPSCTCT--------GTVV 726
```

RESULT 13
PCT-US91-09055-2
; Sequence 2, Application PC/TUS9109055
; GENERAL INFORMATION:
; APPLICANT: Rothberg, Jonathan Marc and Artavanis-Tsakonas, Spyridon
; TITLE OF INVENTION: Purified SLIT protein and Sequence Elements Thereof
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University
; ADDRESSEE: Office of Cooperative Research
; STREET: 246 Church Street
; STREET: Suite 401
; CITY: New Haven
; STATE: Connecuticut
; COUNTRY: USA
; ZIP: 06510
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 800 Kb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/09055
; FILING DATE: 19911127
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/624,135
; FILING DATE: 7-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Barth, Richard J.
; REGISTRATION NUMBER: 28,180
; REFERENCE/DOCKET NUMBER: 900964/RSB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 972-1400
; TELEFAX: (212) 370-1622
; TELEX: 236268
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1480 amino acids
; TYPE: AMINO ACIDS
; TOPOLOGY: Linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: signal sequence
; LOCATION: 1 to 36
; IDENTIFICATION METHOD: similarity to other signal
; OTHER INFORMATION: Directs Export
; NAME/KEY: Four Flank-LRR-Flank domains
; LOCATION: 37 to 910
; IDENTIFICATION METHOD: Array of Flank-LRR-Flank

```
;       OTHER INFORMATION:  mediates adhesive events
;       NAME/KEY:  Tandem EGF-like repeats
;       LOCATION:  911 to 1150
;       IDENTIFICATION METHOD:  similarity to tandem EGF-like
;       OTHER INFORMATION:  protein-protein interactions
;       NAME/KEY:  7th EGF-like repeat
;       LOCATION:  1353 to 1393
;       IDENTIFICATION METHOD:  similarity to epidermal growth
;       OTHER INFORMATION:  Involvement in receptor-ligand
;       NAME/KEY:  Alternative splice segment
;       LOCATION:  1394 to 1404
;       IDENTIFICATION METHOD:  experimental
;       OTHER INFORMATION:  developmentally regulated
;       NAME/KEY:  COOH-terminal region
;       LOCATION:  1405 to 1480
;       IDENTIFICATION METHOD:  experimental
PCT-US91-09055-2


  Query Match             10.4%;  Score 350;  DB 5;  Length 1480;
  Best Local Similarity   23.9%;  Pred. No. 5.1e-21;
  Matches  109;  Conservative   65;  Mismatches  152;  Indels  130;  Gaps    9;

Qy      46 TCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEIL 105
           :||  | |::    | |  |:|  ||  :  :|  : | :|  |  :   ||   | |  :
Db     294 SCPHPCRCADGI--VDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKSFSSFRRLRRI 351

Qy     106 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 165
           || |:|  |   | :||  | || |: |::  :|:| |  |  |: | |  | |  |
Db     352 DLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLLNANEISCIRK 411

Qy     166 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM------CNLR--------- 210
           ||  : ||  | | :   :  :: | |: : ::: ::||       ||||
Db     412 DAFRDLHSLSLLSLYD-NNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRWLADYLHKN 470

Qy     211 ------------------------------------------------------------ 210

Db     471 PIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKLSGECRMDSDCPAMCHCEGTTV 530

Qy     211 --------EIPNLTPL--------------IKLD----------ELDLSGNHLSAIRPGS 238
                  |||   ||               :|:|  | |: | | :
Db     531 DCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKRNQLTGIEPNA 590

Qy     239 FQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 298
           |:|  |:|:| : :::|: |    | |  |  :||  | :: :    |  |: |  ::|
Db     591 FEGASHIQELQLGENKIKEISNKMFLGLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNL 650

Qy     299 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVE 358
             ||:|||| : | :  ::  : :  |  ||| |  ::   | :|  : | | :
Db     651 ASNPFNCNCHLAWFAECVRKKSLNGGA--ARCGAPSKVRDVQIKDLPHSEFKCSS----- 703

Qy     359 PPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVM 394
              :||   :  |  | : |        |||:
Db     704 -----ENSEGCLGDGYCPPSCTCT--------GTVV 726


RESULT  14
US-09-182-024A-5
; Sequence 5, Application US/09182024A
; Patent No. 6342370
; GENERAL INFORMATION:
;  APPLICANT:  Connolly, Timothy
;  APPLICANT:  Rajput, Bhanu
;  TITLE OF INVENTION:  Human Slit Polypeptide and Polynucleotides Encoding
;  TITLE OF INVENTION:  Same
;  FILE REFERENCE:  640100-271
;  CURRENT APPLICATION NUMBER: US/09/182,024A
;  CURRENT FILING DATE:  1998-10-29
;  PRIOR APPLICATION NUMBER:  60/063,946
;  PRIOR FILING DATE:  1997-10-31
;  PRIOR APPLICATION NUMBER:  60/096,420
;  PRIOR FILING DATE:  1998-08-13
;  NUMBER OF SEQ ID NOS:  5
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 1480
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-182-024A-5


  Query Match             10.3%;  Score 347;  DB 4;  Length 1480;
  Best Local Similarity   23.9%;  Pred. No. 9.2e-21;
  Matches  109;  Conservative   65;  Mismatches  152;  Indels  130;  Gaps    9;

Qy      46 TCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEIL 105
           :||  | |::    | |  |:|  ||  :  :|  : | :|  |  :   ||   | |  :
Db     294 SCPHPCRCADGI--VDCREKSLTSVPVTLPDDTTDVLLEQNFITELPPKSFSSFRRLRRI 351

Qy     106 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 165
           || |:|  |   | :||  | || |: |::  :|:| |  |  |: | |  | |  |
Db     352 DLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLLNANEISCIRK 411

Qy     166 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM------CNLR--------- 210
           ||  : ||  | | :   :  :: | |: : ::: ::||       ||||
Db     412 DAFRDLHSLSLLSLYD-NNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRWLADYLHKN 470

Qy     211 ------------------------------------------------------------ 210

Db     471 PIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKLSGECRMDSDCPAMCHCEGTTV 530

Qy     211 --------EIPNLTPL--------------IKLD----------ELDLSGNHLSAIRPGS 238
                  |||   ||               :|:|  | |: | | :
Db     531 DCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKRNQLTGIEPNA 590

Qy     239 FQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 298
           |:|  |:|:| : :::|: |    | |  |  :||  | :: :    |  |: |  ::|
Db     591 FEGASHIQELQLGENKIKEISNKMFLGLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNL 650

Qy     299 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVE 358
             ||:|||| : | :  ::  : :  |  ||| |  ::   | :|  : | | :
Db     651 ASNPFNCNCHLAWFAECVRKKSLNGGA--ARCGAPSKVRDVQIKDLPHSEFKCSS----- 703

Qy     359 PPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVM 394
              :||   :  |  | : |        |||:
Db     704 -----ENSEGCLGDGYCPPSCTCT--------GTVV 726


RESULT  15
US-09-063-950-5
; Sequence 5, Application US/09063950C
; Patent No. 6225085
; GENERAL INFORMATION:
;  APPLICANT: Holtzman, Douglas A.
;  TITLE OF INVENTION: NOVEL LRSG PROTEIN AND NUCLEIC ACID MOLECULES AND USES
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: MEI-019
;  CURRENT APPLICATION NUMBER: US/09/063,950C
;  CURRENT FILING DATE: 1998-04-21
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 605
;   TYPE: PRT
;   ORGANISM: Papio hamadryas
US-09-063-950-5


  Query Match             10.0%;  Score 335;  DB 4;  Length 605;
  Best Local Similarity   20.8%;  Pred. No. 2.5e-20;
  Matches  125;  Conservative   64;  Mismatches  125;  Indels  286;  Gaps   13;

Qy      31 LALQLLVVA----------------GLVRAQTCPSVCSCS-----NQFSKVICVRKNLR 68
           ||| ||:::                |      ||: |:||     |: | | |  :||
Db       8 LALALLLLSWVALGPRSLEGAEPGTPGEAEGPACPATCACSYDDEVNELS-VFCSSRNLT 66
```

```
Qy      69 EVPDGISTNTRL---------------------------------------------- 80
           :||||   |:
Db      67 RLPDGIPGGTQALWLDSNNLSSIPPAAFRNLSSLAFLNLQGGQLGSLEPQALLGLENLCH 126

Qy      81 LNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIP 140
           |:|  ||:: : | :| :   | :| || | :  :| | | || ||  | |  | |  :|
Db     127 LHLERNQLRSLAVGTFAYTPALALLGLSNNRLSRLEDGLFEGLGNLWDLNLGWNSLAVLP 186

Qy     141 NGAF----------------VY--------LSKLKELWLRNNPIESIPSYAFNRIPSLRR 176
           : ||                |         |::|:|| |  | : :| :  | ::| |::
Db     187 DAAFRGLGGLRELVLAGNRLAYLQPALFSGLAELRELDLSRNALRAIKANVFAQLPRLQK 246

Qy     177 L---------------------------------------------------------- 177
           |
Db     247 LYLDRNLIAAVAPGAFLGLKALRWLDLSHNRVAGLLEDTFPGLLGLRVLRLSHNAIASLR 306

Qy     178 ----------------------------DLGELK-------RLSYISEGAFEGLSNLRY 201
                                        ||:|:       :|  :  ||| ||:|:
Db     307 PRTFEDLHFLEELQLGHNRIRQLAERSFEGLGQLEVLTLDHNQLQEVKVGAFLGLTNVAV 366

Qy     202 LNLAMCNLREIPN--LTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIE 259
           :||:   || :|      | ||  | | |: |  ||| :| ||  |::|::  : :  ||
Db     367 MNLSGNCLRNLPEQVFRGLGKLHSLHLEGSCLGRIRPHTFAGLSGLRRLFLKDNGLVGIE 426

Qy     260 RNAFDNLQSLVEINLAHNNLTLLPHDLFTPL---------------------------- 290
             :  |  |:|::|  | || ||| ||  |
Db     427 EQSLWGLAELLELDLTSNQLTHLPHQLFQGLGKLEYLLLSHNRLAELPADALGPLQRAFW 486

Qy     291 ----HH-----------------------------------LERIHLHHNPWNCNCDILW 311
               |:                                  |||: |  |||:|:| :
Db     487 LDVSHNRLEALPGSLLASLGRLRYLNLRNNSLRTFTPQPPGLERLWLEGNPWDCSCPLKA 546

Qy     312 LSWWIKDMAPSNTACCAR---------------------CNTPPNLKGRYIGELDQNYF 349
           |    :| |  | :   |                     | :|| : |  : :| : :|
Db     547 L----RDFALQNPSAVPRFVQAICEGDDCQPPVYTYNNITCASPPEVAGLDLRDLGEAHF 602
```

Search completed: August 26, 2002, 15:41:22
Job time: 225 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 2002, 15:41:53 ; Search time 31.57 Seconds
(without alignments)
1681.764 Million cell updates/sec

Title: US-09-905-056-292_COPY_45_522
Perfect score: 2524
Sequence: 1 QTCPSVCSCSNQFSKVICVR.....KTFTIPVTDINSGIPGIDEV 478

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2524	100.0	640	20	AAY08100	Human PRO331 prote
2	2524	100.0	640	20	AAW85722	Novel protein (Clo
3	2524	100.0	640	20	AAY13394	Amino acid sequenc
4	2524	100.0	640	21	AAB24407	Human PRO331 prote
5	2524	100.0	640	21	AAY70673	Human PRO331 prote
6	2524	100.0	640	22	AAU12355	Human PRO331 polyp
7	2524	100.0	640	22	AAU00826	Human immune respo
8	2524	100.0	640	22	AAB80262	Human PRO331 prote
9	2524	100.0	640	22	AAB65292	Human PRO331 prote
10	2524	100.0	640	22	AAB53089	Human angiogenesis
11	1793.5	71.1	713	22	AAE13006	Human leucine-rich
12	1581.5	62.7	653	20	AAY28806	cc359_4 secreted p
13	1581.5	62.7	653	21	AAB24073	Human PRO1111 prot
14	1581.5	62.7	653	21	AAY66694	Membrane-bound pro
15	1581.5	62.7	653	22	AAE09438	Human sbgPRO331a p
16	1581.5	62.7	653	22	AAU12390	Human PRO1111 poly
17	1581.5	62.7	653	22	AAB65217	Human PRO1111 (UNQ
18	1561.5	61.9	448	21	AAB23044	Human SLIT protein
19	1560.5	61.8	606	22	AAU18035	Human immunoglobul
20	1548.5	61.4	590	21	AAB23034	Human SLIT protein
21	1548.5	61.4	694	21	AAB23033	Human SLIT protein
22	1356.5	53.7	553	22	ABB10349	Human cDNA SEQ ID
23	971	38.5	302	21	AAY87066	Human secreted pro
24	971	38.5	302	22	AAE06043	Human gene 3 encod
25	886.5	35.1	441	21	AAB43091	Human ORFX ORF2855
26	845.5	33.5	281	22	ABB10522	Human cDNA SEQ ID
27	845.5	33.5	281	22	AAU18084	Human immunoglobul
28	685	27.1	224	21	AAY87135	Human secreted pro
29	685	27.1	224	22	AAE06112	Human gene 3 encod
30	568.5	22.5	614	20	AAW84596	Amino acid sequenc
31	565.5	22.4	579	22	AAE06804	Mature human neuro
32	565.5	22.4	606	22	AAE06799	Human neuronal gui
33	565.5	22.4	620	22	AAB74705	Human membrane ass
34	556.5	22.0	620	20	AAY13357	Amino acid sequenc
35	556.5	22.0	620	22	AAU12333	Human PRO227 polyp
36	556.5	22.0	620	22	AAB80225	Human PRO227 prote
37	522.5	20.7	592	22	AAE09437	Human sbgTango79a
38	477.5	18.9	548	22	AAB31161	Amino acid sequenc
39	453	17.9	766	22	AAG67505	Amino acid sequenc
40	443.5	17.6	1021	22	ABB60162	Drosophila melanog
41	430.5	17.1	713	20	AAY13385	Amino acid sequenc
42	430.5	17.1	713	22	AAB80253	Human PRO293 prote
43	422.5	16.7	707	20	AAY02379	Polypeptide identi
44	421	16.7	705	21	AAB42513	Human ORFX ORF2277
45	421	16.7	705	22	AAM78823	Human protein SEQ

ALIGNMENTS

```
RESULT 1
AAY08100
ID   AAY08100 standard; Protein; 640 AA.
XX
AC   AAY08100;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Human PRO331 protein.
XX
KW   Inflammatory cell infiltration; immune response; T cell proliferation;
KW   anti-inflammatory; anti-autoimmune; anti-diabetic; spondyloarthropathy;
KW   T cell-mediated disease; spondyloarthropathy; sclerosis; renal disease;
KW   inflammatory myopathy; hemolytic anemia; thrombocytopenia; thyroiditis;
KW   diabetes mellitus; demyelinating polyneuropathy; Guillain-Barre syndrome;
KW   multiple sclerosis; polyneuropathy; hepatitis; cirrhosis; enteropathy;
KW   sclerosing cholangitis; inflammatory bowel disease; Whipple's disease;
KW   skin disease; dermatitis; psoriasis; asthma; allergic rhinitis; tumor;
KW   food hypersensitivity; urticaria; eosinophilic pneumonia; transplant;
KW   idiopathic pulmonary fibrosis; graft rejection; PRO245; human;
KW   PRO331.
XX
OS   Homo sapiens.
XX
PN   WO9914241-A2.
XX
PD   25-MAR-1999.
XX
PF   17-SEP-1998;   98WO-US19437.
XX
PR   17-SEP-1997;   97US-0059119.
PR   18-SEP-1997;   97US-0059263.
PR   28-OCT-1997;   97US-0063550.
PR   12-NOV-1997;   97US-0065186.
```

PR 21-NOV-1997; 97US-0066364.
PR 24-NOV-1997; 97US-0066770.
PR 04-JUN-1998; 98US-0088026.
XX
PA (GETH) GENENTECH INC.
XX
PI Fong S, Goddard A, Gurney AL, Tumas D, Wood WI;
XX
DR WPI; 1999-229499/19.
DR N-PSDB; AAX37738.
XX
PT Composition containing novel polypeptide PRO245, its agonist or
PT antagonist -
XX
PS Example 1; Fig 33; 177pp; English.
XX
CC This invention describes a novel composition containing (apart from a
CC carrier or excipient), a novel PRO245 polypeptide (I), its agonist or
CC antagonist, or their fragments, for modulating: (i) infiltration of
CC inflammatory cells into tissue; (ii) an immune response; or (iii) T cell
CC proliferation. The composition increases or decreases any of the effects
CC (i)-(iii). The products of the invention have anti-inflammatory,
CC anti-autoimmune and anti-diabetic activity. (I), and its (ant)agonists
CC and their fragments, are used to treat immune-related diseases,
CC particularly T cell-mediated diseases. The diseases treated include
CC systemic lupus erythematosus, rheumatoid arthritis, juvenile chronic
CC arthritis, spondyloarthropathies, systemic sclerosis (scleroderma),
CC idiopathic inflammatory myopathies (dermatomyositis, polymyositis),
CC Sjogren's syndrome, systemic vasculitis, sarcoidosis, autoimmune
CC hemolytic anemia (immune pancytopenia, paroxysmal nocturnal
CC hemoglobinuria), autoimmune thrombocytopenia (idiopathic thrombocytopenic
CC purpura immune-mediated thrombocytopenia), thyroiditis (Grave's disease,
CC Hashimoto's thyroiditis, juvenile lymphocytic thyroiditis, atrophic
CC thyroiditis), diabetes mellitus, immune-mediated renal disease
CC (glomerulonephritis, tubulointerstitial nephritis), multiple sclerosis,
CC idiopathic demyelinating polyneuropathy, Guillain-Barre syndrome, chronic
CC inflammatory demyelinating polyneuropathy, infectious hepatitis
CC (hepatitis A, B, C, D, E and other non-hepatotropic viruses), autoimmune
CC chronic active hepatitis, primary biliary cirrhosis, granulomatous
CC hepatitis, and sclerosing cholangitis, inflammatory bowel disease
CC (ulcerative colitis: Crohn's disease), gluten-sensitive enteropathy, and
CC Whipple's disease. Autoimmune or immune-mediated skin diseases including
CC bullous skin diseases, erythema multiforme,contact dermatitis, psoriasis,
CC asthma, allergic rhinitis, atopic dermatitis, food hypersensitivity,
CC urticaria, eosinophilic pneumonia, idiopathic pulmonary fibrosis,
CC hypersensitivity pneumonitis, and transplantation associated diseases
CC (graft rejection, and graft-versus-host-disease). (I), its (ant)agonists
CC or fragment can also be used as an adjuvant in treatment of tumors.
CC Antibodies against (I) can also be used for diagnosing such diseases.
CC This sequence represents the human PRO331 protein which is described in
CC the method of the invention.
XX
SQ Sequence 640 AA;

Query Match 100.0%; Score 2524; DB 20; Length 640;
Best Local Similarity 100.0%; Pred. No. 2.4e-177;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     45 qtcpsvcscsnqfskvicvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhlei 104

Qy     61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    105 lqlsrnhirtieigafnglanlntlelfdnrlttipngafvylsklkelwlrnnpiesip 164

Qy    121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    165 syafnripslrrldlgelkrlsyisegafeglsnlrylnlamcnlreipnltplikldel 224

Qy    181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    225 dlsgnhlsairpgsfqglmhlqklwmiqsqiqviernafdnlqslveinlahnnltllph 284

Qy    241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGEL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    285 dlftplhhlerihlhhnpwncncdilwlswwikdmapsntaccarcntppnlkgryigel 344

Qy    301 DQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    345 dqnyftcyapviveppadlnvtegmaaelkcrastsltsvswitpngtvmthgaykvria 404

Qy    361 VLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    405 vlsdgtlnftnvtvqdtgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmep 464

Qy    421 SQDEARTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEV 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    465 sqdearttdnnvgptpvvdwettnvttsltpqstrstektftipvtdinsgipgidev 522
```

RESULT 2
AAW85722
ID AAW85722 standard; Protein; 640 AA.
XX
AC AAW85722;
XX
DT 27-SEP-1999 (first entry)
XX
DE Novel protein (Clone AS209_1).
XX
KW Polynucleotide; protein; nutrition; cytokine; cell proliferation;
KW cell differentiation; immunostimulation; immunosuppression;
KW haematopoiesis regulation; tissue growth; activin; inhibin;
KW chemotaxis; chemokinesis; haemostasis; thrombolysis; receptor;
KW ligand; anti-inflammatory; tumour suppression; gene therapy.
XX
OS Homo sapiens.
XX
PN WO9920644-A1.
XX
PD 29-APR-1999.
XX
PF 16-OCT-1998; 98WO-US22034.
XX
PR 18-OCT-1997; 97US-0955557.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Agostino MJ, Bowman MR, Evans C, Jacobs K, Lavallie ER;
PI McCoy JM, Merberg D, Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1999-288272/24.
DR N-PSDB; AAX08687.
XX
PT New polynucleotides encoding secreted human proteins
XX
PS Claim 26; Page 109-111; 136pp; English.
XX
CC The new human secreted proteins are encoded by polynucleotides
CC obtained from human placenta, adult testes, fetal kidney, fetal
CC brain, adult brain, adult brain and adult blood cDNA libraries.
CC The polynucleotides and proteins are predicted to have biological
CC activities which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals. Suggested
CC activities include nutritional activity, cytokine and cell
CC proliferation/differentiation activity, immune stimulating (e.g. as
CC vaccines) or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cadherin/tumour invasion suppressor activity, and tumour inhibition
CC activity. The polynucleotides are also stated to be useful for gene
CC therapy. The sequences identified by a secretory leader
CC sequence motif in the polynucleotide and it is thought that the

```
CC   encoded proteins have biological activity by virtue of their secreted
CC   nature. This polypeptide was encoded by a clone designated AS209_1
CC   (See AAX08687).
XX
SQ   Sequence   640 AA;


  Query Match             100.0%;  Score 2524;  DB 20;  Length 640;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-177;
  Matches  478;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45 qtcpsvcscsnqfskvicvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhlei 104

Qy      61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     105 lqlsrnhirtieigafnglanlntlelfdnrlttipngafvylsklkelwlrnnpiesip 164

Qy     121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165 syafnripslrrldlgelkrlsyisegafeglsnlrylnlamcnlreipnltplikldel 224

Qy     181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     225 dlsgnhlsairpgsfqglmhlqklwmiqsqiqviernafdnlqslveinlahnnltllph 284

Qy     241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGEL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     285 dlftplhhlerihlhhnpwncncdilwlswwikdmapsntaccarcntppnlkgryigel 344

Qy     301 DQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     345 dqnyftcyapviveppadlnvtegmaaelkcrastsltsvswitpngtvmthgaykvria 404

Qy     361 VLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     405 vlsdgtlnftnvtvqdtgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmep 464

Qy     421 SQDEARTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEV 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     465 sqdearttdnnvgptpvvdwettnvttsltpqstrstektftipvtdinsgipgidev 522


RESULT   3
AAY13394
ID   AAY13394 standard; Protein; 640 AA.
XX
AC   AAY13394;
XX
DT   25-JUN-1999  (first entry)
XX
DE   Amino acid sequence of protein PRO331.
XX
KW   Secreted protein; transmembrane protein; human; enterocolitis;
KW   Zollinger-Ellison syndrome; gastrointestinal ulceration;
KW   congenital microvillus atrophy; skin disease; cell growth;
KW   abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
KW   Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
KW   fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;
KW   anti-thrombotic; wound healing; tissue repair.
XX
OS   Homo sapiens.
XX
PN   WO9914328-A2.
XX
PD   25-MAR-1999.
XX
PF   16-SEP-1998;   98WO-US19330.
XX
PR   25-NOV-1997;   97US-0066840.
PR   17-SEP-1997;   97US-0059113.
PR   17-SEP-1997;   97US-0059115.
PR   17-SEP-1997;   97US-0059117.
PR   17-SEP-1997;   97US-0059119.
PR   17-SEP-1997;   97US-0059121.
PR   17-SEP-1997;   97US-0059122.
PR   17-SEP-1997;   97US-0059184.
PR   18-SEP-1997;   97US-0059263.
PR   18-SEP-1997;   97US-0059266.
PR   15-OCT-1997;   97US-0062125.
PR   17-OCT-1997;   97US-0062285.
PR   17-OCT-1997;   97US-0062287.
PR   21-OCT-1997;   97US-0063486.
PR   24-OCT-1997;   97US-0062814.
PR   24-OCT-1997;   97US-0062816.
PR   24-OCT-1997;   97US-0063045.
PR   24-OCT-1997;   97US-0063120.
PR   24-OCT-1997;   97US-0063121.
PR   24-OCT-1997;   97US-0063127.
PR   24-OCT-1997;   97US-0063128.
PR   27-OCT-1997;   97US-0063329.
PR   27-OCT-1997;   97US-0063327.
PR   28-OCT-1997;   97US-0063541.
PR   28-OCT-1997;   97US-0063542.
PR   28-OCT-1997;   97US-0063544.
PR   28-OCT-1997;   97US-0063549.
PR   28-OCT-1997;   97US-0063550.
PR   28-OCT-1997;   97US-0063564.
PR   29-OCT-1997;   97US-0063435.
PR   29-OCT-1997;   97US-0063704.
PR   29-OCT-1997;   97US-0063732.
PR   29-OCT-1997;   97US-0063738.
PR   29-OCT-1997;   97US-0063734.
PR   29-OCT-1997;   97US-0064215.
PR   29-OCT-1997;   97US-0063735.
PR   31-OCT-1997;   97US-0063870.
PR   31-OCT-1997;   97US-0064103.
PR   03-NOV-1997;   97US-0064248.
PR   07-NOV-1997;   97US-0064809.
PR   12-NOV-1997;   97US-0065186.
PR   17-NOV-1997;   97US-0065846.
PR   18-NOV-1997;   97US-0065693.
PR   21-NOV-1997;   97US-0066120.
PR   21-NOV-1997;   97US-0066364.
PR   24-NOV-1997;   97US-0066772.
PR   24-NOV-1997;   97US-0066466.
PR   24-NOV-1997;   97US-0066770.
PR   24-NOV-1997;   97US-0066511.
PR   24-NOV-1997;   97US-0066453.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Chen J,  Goddard A,  Gurney AL,  Pennica D,  Wood WI,  Yuan J;
XX
DR   WPI; 1999-229533/19.
DR   N-PSDB; AAX52265.
XX
PT   New isolated human genes and polypeptides used in, e.g. treatment of
PT   gastrointestinal ulceration
XX
PS   Claim 12; Fig 104; 320pp; English.
XX
CC   AAY13344-403 represent secreted and transmembrane human proteins.
CC   The cDNA sequences are obtained from cDNA libraries, prepared from
CC   fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
CC   The encoded polypeptides have specific uses based on their homology to
CC   known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
CC   associated with the preservation and maintenance of gastrointestinal
CC   mucosa and the repair of acute and chronic mucosal lesions
CC   (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
CC   ulceration and congenital microvillus atrophy), skin diseases associated
CC   with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
CC   cancers such as lung squamous cell carcinoma of the vulva and gliomas),
CC   potent effects on cell growth and development, diseases related to
CC   growth or survival of nerve cells including Parkinson's disease,
```

```
CC   Alzheimer's disease, ALS, neuropathies or cancer. PRO265 can be used as
CC   for fibromodulin, e.g. for reducing dermal scarring.  PRO264 can be used
CC   as a target for anti-tumor drugs. PRO533 may be used in the treatment
CC   of Usher Syndrome or Atrophia areata; PRO269 can be used as an
CC   anti-thrombotic agent; PRO287 polypeptides and portions may have
CC   therapeutic applications in wound healing and tissue repair; PRO317 can
CC   be used for treating problems of the kidney, uterus, endometrium, blood
CC   vessels, or related tissue, e.g. in the heart of genital tract.
XX
SQ   Sequence   640 AA;


  Query Match            100.0%;  Score 2524;  DB 20;  Length 640;
  Best Local Similarity  100.0%;  Pred. No. 2.4e-177;
  Matches  478;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     45 qtcpsvcscsnqfskvicvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhlei 104

Qy     61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    105 lqlsrnhirtieigafnglanlntlelfdnrlttipngafvylsklkelwlrnnpiesip 164

Qy    121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    165 syafnripslrrldlgelkrlsyisegafeglsnlrylnlamcnlreipnltplikldel 224

Qy    181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    225 dlsgnhlsairpgsfqglmhlqklwmiqsqiqviernafdnlqslveinlahnnltllph 284

Qy    241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGEL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    285 dlftplhhlerihlhhnpwncncdilwlswwikdmapsntaccarcntppnlkgryigel 344

Qy    301 DQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    345 dqnyftcyapviveppadlnvtegmaaelkcrastsltsvswitpngtvmthgaykvria 404

Qy    361 VLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    405 vlsdgtlnftnvtvqdtgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmep 464

Qy    421 SQDEARTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEV 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    465 sqdearttdnnvgptpvvdwettnvttsltpqstrstektftipvtdinsgipgidev 522


RESULT   4
AAB24407
ID   AAB24407 standard; Protein; 640 AA.
XX
AC   AAB24407;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Human PRO331 protein sequence SEQ ID NO:107.
XX
KW   Human; PRO; promotion; inhibition; angiogenesis; cardiovascularisation;
KW   diagnosis; trauma; wound; cancer; atherosclerosis; cardiac hypertrophy;
KW   angiogenic; proliferative; cardiant; cardiovascular; antiatherosclerotic;
KW   cytostatic; gene therapy; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200032221-A2.
XX
PD   08-JUN-2000.
XX
PF   30-NOV-1999;   99WO-US28313.
XX
PR   01-DEC-1998;   98WO-US25108.
PR   16-DEC-1998;   98US-0112850.
PR   12-JAN-1999;   99US-0115554.
PR   08-MAR-1999;   99WO-US05028.
PR   12-MAR-1999;   99US-0123957.
PR   28-APR-1999;   99US-0131445.
PR   14-MAY-1999;   99US-0134287.
PR   02-JUN-1999;   99WO-US12252.
PR   23-JUN-1999;   99US-0141037.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   01-SEP-1999;   99WO-US20111.
PR   08-SEP-1999;   99WO-US20594.
PR   13-SEP-1999;   99WO-US20944.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   05-OCT-1999;   99WO-US23089.
PR   29-OCT-1999;   99US-0162506.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Ferrara N,  Gerber H,  Hillan KJ,  Goddard A;
PI   Godowski PJ,  Gurney AL,  Klein RD,  Kuo SS,  Paoni NF,  Smith V;
PI   Watanabe CK,  Williams PM,  Wood WI;
XX
DR   WPI; 2000-412154/35.
DR   N-PSDB; AAA77596.
XX
PT   Nucleic acids encoding PRO polypeptides useful for preventing,
PT   diagnosing and treating diagnosing a cardiovascular, endothelial or
PT   angiogenic disorders in mammals -
XX
PS   Claim 72; Fig 40; 315pp; English.
XX
CC   The present invention describes nucleic acids encoding PRO polypeptides
CC   useful for preventing, diagnosing and treating diagnosing a
CC   cardiovascular, endothelial or angiogenic disorder in mammals by
CC   modulating cell proliferation, angiogenesis and cardiovascularisation,
CC   and for identifying agonists and antagonists of these processes. The
CC   nucleic acids and the proteins they encode may be used in the
CC   prevention, treatment and diagnosis of diseases associated with
CC   inappropriate PRO expression such as cardiovascular, endothelial or
CC   angiogenic disorders in mammals (e.g. atherosclerosis, cancers and
CC   cardiac hypertrophy). For example, the nucleic acids (NCs) and vectors
CC   containing them and the PRO polypeptide may be used to treat disorders
CC   associated with decreased PRO expression. AAA77510 to AAA77721 and
CC   AAB24388 to AAB24435 represent nucleotide and protein sequences used in
CC   the exemplification of the present invention.
XX
SQ   Sequence   640 AA;


  Query Match            100.0%;  Score 2524;  DB 21;  Length 640;
  Best Local Similarity  100.0%;  Pred. No. 2.4e-177;
  Matches  478;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     45 qtcpsvcscsnqfskvicvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhlei 104

Qy     61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    105 lqlsrnhirtieigafnglanlntlelfdnrlttipngafvylsklkelwlrnnpiesip 164

Qy    121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    165 syafnripslrrldlgelkrlsyisegafeglsnlrylnlamcnlreipnltplikldel 224

Qy    181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    225 dlsgnhlsairpgsfqglmhlqklwmiqsqiqviernafdnlqslveinlahnnltllph 284

Qy    241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGEL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    285 dlftplhhlerihlhhnpwncncdilwlswwikdmapsntaccarcntppnlkgryigel 344

Qy    301 DQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    345 dqnyftcyapviveppadlnvtegmaaelkcrastsltsvswitpngtvmthgaykvria 404

Qy    361 VLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    405 vlsdgtlnftnvtvqdtgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmep 464

Qy    421 SQDEARTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEV 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    465 sqdearttdnnvgptpvvdwettnvttsltpqstrstektftipvtdinsgipgidev 522


RESULT   5
AAY70673
ID   AAY70673 standard; Protein; 640 AA.
XX
AC   AAY70673;
XX
DT   18-JUL-2000  (first entry)
XX
DE   Human PRO331 protein.
XX
KW   PRO331; UNQ292; dermatological; immunosuppressive; antiinflammatory;
KW   immunostimulant; antiasthmatic; antirheumatic; antiarthritic; virucide;
KW   antiallergic; haemostatic; hepatotropic; antidiabetic; antianaemic;
KW   nephrotropic; neuroprotective; anticoagulant; immunological disorder;
KW   lung; pneumonia; skin; psoriasis; kidney; glomerulonephritis; arthritis;
KW   spondyloarthropathy; SLE; systemic lupus erythematosis; scleroderma;
KW   idiopathic inflammatory myopathy; anaemia; thrombocytopenia; diabetes;
KW   thyroiditis; Grave's disease; demyelinating disease; multiple sclerosis;
KW   Crohn's disease; hepatobiliary disease; hepatitis; asthma; human;
KW   graft-versus-host-disease.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   40..46
FT                   /note= "N-myristoylation site"
FT   Modified-site   73..79
FT                   /note= "N-myristoylation site"
FT   Modified-site   118..124
FT                   /note= "N-myristoylation site"
FT   Modified-site   183..187
FT                   /note= "cAMP and cGMP-dependent protein kinase
FT                   phosphorylation site"
FT   Modified-site   191..197
FT                   /note= "N-myristoylation site"
FT   Modified-site   228..234
FT                   /note= "N-myristoylation site"
FT   Modified-site   237..243
FT                   /note= "N-myristoylation site"
FT   Modified-site   268..272
FT                   /note= "Casein Kinase II phosphorylation site"
FT   Modified-site   278..282
FT                   /note= "N-glycosylation site"
FT   Modified-site   364..368
FT                   /note= "N-glycosylation site"
FT   Modified-site   390..394
FT                   /note= "N-glycosylation site"
FT   Modified-site   391..397
FT                   /note= "N-myristoylation site"
FT   Modified-site   412..416
FT                   /note= "N-glycosylation site"
FT   Modified-site   415..419
FT                   /note= "N-glycosylation site"
FT   Modified-site   417..421
FT                   /note= "Casein Kinase II phosphorylation site"
FT   Modified-site   422..428
FT                   /note= "N-myristoylation site"
FT   Modified-site   433..439
FT                   /note= "N-myristoylation site"
FT   Modified-site   434..438
FT                   /note= "N-glycosylation site"
FT   Modified-site   442..446
FT                   /note= "N-glycosylation site"
FT   Modified-site   465..469
FT                   /note= "Casein Kinase II phosphorylation site"
FT   Modified-site   488..492
FT                   /note= "N-glycosylation site"
FT   Modified-site   531..537
FT                   /note= "N-myristoylation site"
FT   Modified-site   579..583
FT                   /note= "Casein Kinase II phosphorylation site"
FT   Modified-site   606..610
FT                   /note= "N-glycosylation site"
FT   Modified-site   620..624
FT                   /note= "Casein Kinase II phosphorylation site"
XX
PN   WO200015797-A2.
XX
PD   23-MAR-2000.
XX
PF   15-SEP-1999;   99WO-US21547.
XX
PR   17-SEP-1998;   98US-0100858.
PR   17-SEP-1998;   98WO-US19437.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Fong S,  Goddard A,  Gurney AL,  Tumas D,  Wood WI;
XX
DR   WPI; 2000-271435/23.
DR   N-PSDB; AAZ52207.
XX
PT   Composition for treatment and diagnosis of immune related diseases e.g.
PT   Grave's disease comprises a PRO245, PRO217, PRO301, PRO266, PRO335,
PT   PRO331 or PRO326 polypeptide or its agonists or antagonists (preferably
PT   antibodies) -
XX
PS   Example 1; Fig 14; 201pp; English.
XX
CC   The present sequence is the human protein PRO331, encoded by UNQ292 cDNA,
CC   designated as clone DNA40981. It is isolated from human foetal brain
CC   tissue. Portions of PRO331 has homology to the LIG-1 protein.
CC   It enhances or suppresses the infiltration of inflammatory cells into
CC   tissues, proliferation of T-lymphocytes and modulates the immune
CC   response. This sequence is useful for treatment of immune related
CC   disorders, like SLE, rheumatoid/juvenile arthritis, spondyloarthropathy,
CC   systemic sclerosis (scleroderma), idiopathic inflammatory myopathies such
CC   as dermatomyositis, Sjogren's syndrome, systemic vasculitis, sarcoidosis,
CC   autoimmune haemolytic anaemia, thrombocytopenia, thyroiditis e.g. Grave's
CC   disease, diabetes mellitus, immune-mediated renal disease e.g.
CC   glomerulonephritis, demyelinating diseases such as multiple sclerosis and
CC   Guillain-Barre syndrome, hepatobiliary diseases like hepatitis and
CC   primary biliary cirrhosis, inflammatory and fibrotic lung diseases such
CC   as inflammatory bowel disease (e.g. Crohn's disease), autoimmune or
CC   immune-mediated skin diseases such as psoriasis, allergies like asthma,
CC   immunological diseases of the lungs such as eosinophilic pneumonia and
CC   transplantation associated diseases such as graft-versus-host-disease.
XX
SQ   Sequence   640 AA;


  Query Match            100.0%;  Score 2524;  DB 21;  Length 640;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-177;
  Matches  478;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     45 qtcpsvcscsnqfskvicvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhlei 104

Qy     61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    105 lqlsrnhirtieigafnglanlntlelfdnrlttipngafvylsklkelwlrnnpiesip 164

Qy    121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    165 syafnripslrrldlgelkrlsyisegafeglsnlrylnlamcnlreipnltplikldel 224

Qy    181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    225 dlsgnhlsairpgsfqglmhlqklwmiqsqiqviernafdnlqslveinlahnnltllph 284

Qy    241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGEL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    285 dlftplhhlerihlhhnpwncncdilwlswwikdmapsntaccarcntppnlkgryigel 344

Qy    301 DQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    345 dqnyftcyapviveppadlnvtegmaaelkcrastsltsvswitpngtvmthgaykvria 404

Qy    361 VLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    405 vlsdgtlnftnvtvqdtgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmep 464

Qy    421 SQDEARTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEV 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db    465 sqdearttdnnvgptpvvdwettnvttsltpqstrstektftipvtdinsgipgidev 522


RESULT   6
AAU12355
ID   AAU12355 standard; Protein; 640 AA.
XX
AC   AAU12355;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human PRO331 polypeptide sequence.
XX
KW   Human secretory and transmembrane; PRO; mammalian; cancer; lung;
KW   breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW   cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW   adipocyte; A-peptide; factor VIIA; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200140466-A2.
XX
PD   07-JUN-2001.
XX
PF   01-DEC-2000; 2000WO-US32678.
XX
PR   01-DEC-1999;   99WO-US28301.
PR   01-DEC-1999;   99WO-US28634.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28564.
PR   02-DEC-1999;   99WO-US28565.
PR   09-DEC-1999;   99US-0170262.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   20-DEC-1999;   99WO-US30999.
PR   30-DEC-1999;   99WO-US31243.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
PR   11-FEB-2000; 2000WO-US03565.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   24-FEB-2000; 2000WO-US04914.
PR   24-FEB-2000; 2000WO-US05004.
PR   01-MAR-2000; 2000WO-US05601.
PR   20-MAR-2000; 2000WO-US07377.
PR   21-MAR-2000; 2000WO-US07532.
PR   30-MAR-2000; 2000WO-US08439.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   10-NOV-2000; 2000WO-US30873.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Beresini M,  Deforge L,  Desnoyers L,  Filvaroff E,  Gao W;
PI   Gerritsen ME,  Goddard A,  Godowski PJ,  Gurney AL,  Sherwood S;
PI   Smith V,  Stewart TA,  Tumas D,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2001-408281/43.
DR   N-PSDB; AAS21427.
XX
PT   Isolated, secretory and transmembrane PRO polypeptide used to detect
PT   other PRO polypeptides, link bioactive molecules to cells expressing
PT   PRO polypeptides, and detect the presence of mammalian tumours e.g.
PT   lung, breast, prostate, cervical  -
XX
PS   Claim 12; Fig 368; 813pp; English.
XX
CC   AAU12172-AAU12446 represent novel human secretory and transmembrane
CC   PRO polypeptides. The PRO polypeptides are useful to detect other
CC   PRO polypeptides, to link bioactive molecules to cells expressing
CC   PRO polypeptides, to modulate biological activities of cells expressing
CC   PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC   breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC   polypeptide expression in a cell sample to that in a control sample.
CC   Some of the 275 sequences are also useful to stimulate the release of
CC   tumour necrosis factor-alpha (TNF-alpha) from human blood, the
CC   proliferation or differentiation of chondrocytes, the proliferation or
CC   gene expression in pericyte cells, the release of proteoglycans from
CC   cartilage, the proliferation of inner ear utricular supporting cells or
CC   of T-lymphocytes, the release of a cytokine from peripheral blood
CC   monocytes (PBMCs), or the proliferation of endothelial cells. Some of
CC   the PRO polypeptides may modulate glucose or free fatty acid uptake by
CC   skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
CC   to factor VIIA. The PRO polypeptides can be used in assays to identify
CC   molecules involved in binding interactions. The polynucleotides encoding
CC   PRO polypeptides can be used to generate probes, antisense RNA/DNA,
CC   transgenic or knock out animals and can be used in gene therapy.
XX
SQ   Sequence   640 AA;


  Query Match           100.0%;  Score 2524;  DB 22;  Length 640;
  Best Local Similarity  100.0%;  Pred. No. 2.4e-177;
  Matches  478;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     45 qtcpsvcscsnqfskvicvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhlei 104

Qy     61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    105 lqlsrnhirtieigafnglanlntlelfdnrlttipngafvylsklkelwlrnnpiesip 164

Qy    121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    165 syafnripslrrldlgelkrlsyisegafeglsnlrylnlamcnlreipnltplikldel 224

Qy    181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    225 dlsgnhlsairpgsfqglmhlqklwmiqsqiqviernafdnlqslveinlahnnltllph 284

Qy    241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGEL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    285 dlftplhhlerihlhhnpwncncdilwlswwikdmapsntaccarcntppnlkgryigel 344

Qy    301 DQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    345 dqnyftcyapviveppadlnvtegmaaelkcrastsltsvswitpngtvmthgaykvria 404
```

```
Qy    361 VLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    405 vlsdgtlnftnvtvqdtgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmep 464

Qy    421 SQDEARTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEV 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    465 sqdearttdnnvgptpvvdwettnvttsltpqstrstektftipvtdinsgipgidev 522


RESULT  7
AAU00826
ID   AAU00826 standard; Protein; 640 AA.
XX
AC   AAU00826;
XX
DT   04-JUL-2001  (first entry)
XX
DE   Human immune response protein PRO331 (UNQ292).
XX
KW   Human; PRO331; UNQ292; immune response; osteoarthritis;
KW   systemic lupus erythematosus; rheumatoid arthritis; systemic sclerosis;
KW   juvenile chronic arthritis; spondyloarthropathy; Sjogren's syndrome;
KW   idiopathic inflammatory; myopathy; polymyositis; systemic vasculitis;
KW   sarcoidosis; autoimmune haemolytic anaemia; immune pancytopaenia;
KW   autoimmune thrombocytopaenia; idiopathic thrombocytopaenic purpura;
KW   thyroiditis; Grave's disease; Hashimoto's thyroiditis;
KW   diabetes mellitus; glomerulonephritis; demyelinating disease;
KW   multiple sclerosis; Guillain-Barre syndrome; hepatobiliary disease;
KW   chronic inflammatory demyelinating polyneuropathy; infectious hepatitis;
KW   auto immune chronic active hepatitis; primary biliary cirrhosis;
KW   granulomatous hepatitis; sclerosing cholangitis; ulcerative colitis;
KW   inflammatory bowel disease; Crohn's disease; Whipple's disease;
KW   erythaemia multiforme; psoriasis; asthma; allergic rhinitis; urticaria;
KW   food hypersensitivity; eosinophilic pneumonia; graft rejection;
KW   idiopathic pulmonary fibrosis; graft-versus-host-disease; immunogen;
KW   antibody.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..44
FT                   /label= Signal_peptide
FT   Modified-site   40..46
FT                   /note= "Glycine is N-myristoylated"
FT   Protein         45..640
FT                   /label= Mature_PRO331
FT   Modified-site   73..79
FT                   /note= "Glycine is N-myristoylated"
FT   Modified-site   118..124
FT                   /note= "Glycine at 118 is N-myristoylated"
FT   Region          183..187
FT                   /label= Phosphorylation_site
FT                   /note= "cAMP/cGMP dependent protein kinase
FT                   phosphorylation site"
FT   Modified-site   191..197
FT                   /note= "Glycine at 191 is N-myristoylated"
FT   Modified-site   228..234
FT                   /note= "Glycine is N-myristoylated"
FT   Modified-site   237..243
FT                   /note= "Glycine is N-myristoylated"
FT   Modified-site   278..282
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   364..368
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   390..394
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   391..397
FT                   /note= "Glycine at 391 is N-myristoylated"
FT   Modified-site   412..416
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   415..419
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   422..428
FT                   /note= "Glycine is N-myristoylated"
FT   Modified-site   433..439
FT                   /note= "Glycine is N-myristoylated"
FT   Modified-site   434..438
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   442..446
FT                   /note= "Asn is N-glycosylated"
FT   Modified-site   488..492
FT                   /note= "Asn is N-glycosylated"
FT   Domain          528..543
FT                   /label= Transmembrane_domain
FT   Modified-site   531..537
FT                   /note= "Glycine is N-myristoylated"
FT   Modified-site   606..610
FT                   /note= "Asn is N-glycosylated"
XX
PN   WO200119991-A1.
XX
PD   22-MAR-2001.
XX
PF   20-MAR-2000; 2000WO-US07377.
XX
PR   15-SEP-1999;   99WO-US21547.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Fong S,  Goddard A,  Gurney AL,  Hillan KJ,  Tumas D,  Wood WI;
XX
DR   WPI; 2001-226823/23.
DR   N-PSDB; AAS00162.
XX
PT   Composition for diagnosing and treating immune related diseases, e.g.
PT   rheumatoid arthritis and diabetes mellitus, comprises a PRO
PT   polypeptide, agonist, antagonist or fragment -
XX
PS   Claim 31; Fig 12; 138pp; English.
XX
CC   The sequence represents Human PRO331 (UNQ292), a protein involved in
CC   the immune response. PRO polypeptides, and (ant)agonists to them, are
CC   used in compositions for modulating infiltration of inflammatory cells
CC   into a tissue, modulating an immune response and modulating proliferation
CC   of T-lymphocytes in response to an antigen. Immune related diseases can
CC   be treated with the compositions, such as, systemic lupus erythematosus,
CC   rheumatoid arthritis, osteoarthritis, juvenile chronic arthritis,
CC   spondyloarthopathies, systemic sclerosis, idiopathic inflammatory
CC   myopathies (e.g. polymyositis), Sjogren's syndrome, systemic vasculitis,
CC   sarcoidosis, autoimmune haemolytic anaemia (e.g. immune pancytopaenia),
CC   autoimmune thrombocytopaenia (e.g. idiopathic thrombocytopaenic purpura),
CC   thyroiditis (e.g. Grave's disease, Hashimoto's thyroiditis), diabetes
CC   mellitus, immune-mediated renal disease (e.g. glomerulonephritis),
CC   demyelinating diseases of the central and peripheral nervous systems e.g.
CC   multiple sclerosis or Guillain-Barre syndrome, and chronic inflammatory
CC   demyelinating polyneuropathy, hepatobiliary diseases such as infectious
CC   hepatitis (hepatitis A, B, C, D, E and other non-hepatotropic viruses),
CC   auto immune chronic active hepatitis, primary biliary cirrhosis,
CC   granulomatous hepatitis, and sclerosing cholangitis, inflammatory bowel
CC   disease (ulcerative colitis, Crohn's disease and Whipple's disease),
CC   autoimmune or immune-mediated skin diseases (e.g. erythaemia
CC   multiforme and psoriasis), asthma, allergic rhinitis, urticaria,
CC   food hypersensitivity, immunologic diseases of the lung such as
CC   eosinophilic pneumonias, idiopathic pulmonary fibrosis, transplantation
CC   associated diseases including graft-versus-host-disease and graft
CC   rejection. PRO polypeptides can be used to diagnose immune related
CC   diseases, to identify inhibitors, and to stimulate the proliferation of
CC   T lymphocytes.  Anti-PRO antibodies can be used to detect PRO and
CC   in diagnosis. PRO polypeptides, antibodies and (ant)agonists can be used
CC   in rational drug design.
XX
SQ   Sequence   640 AA;

  Query Match             100.0%;  Score 2524;  DB 22;  Length 640;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-177;
```

```
  Matches  478;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     45 qtcpsvcscsnqfskvicvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhlei 104

Qy     61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    105 lqlsrnhirtieigafnglanlntlelfdnrlttipngafvylsklkelwlrnnpiesip 164

Qy    121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    165 syafnripslrrldlgelkrlsyisegafeglsnlrylnlamcnlreipnltplikldel 224

Qy    181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    225 dlsgnhlsairpgsfqglmhlqklwmiqsqiqviernafdnlqslveinlahnnltllph 284

Qy    241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGEL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    285 dlftplhhlerihlhhnpwncncdilwlswwikdmapsntaccarcntppnlkgryigel 344

Qy    301 DQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    345 dqnyftcyapviveppadlnvtegmaaelkcrastsltsvswitpngtvmthgaykvria 404

Qy    361 VLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    405 vlsdgtlnftnvtvqdtgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmep 464

Qy    421 SQDEARTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEV 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    465 sqdearttdnnvgptpvvdwettnvttsltpqstrstektftipvtdinsgipgidev 522


RESULT  8
AAB80262
ID   AAB80262 standard; Protein; 640 AA.
XX
AC   AAB80262;
XX
DT   24-APR-2001  (first entry)
XX
DE   Human PRO331 protein.
XX
KW   Human; PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;
KW   antiparkinsonian nootropic; neuroprotective; vulnerary; cardiant;
KW   antiangiogenic; vasotropic; antiasthmatic; antirheumatic; cancer;
KW   antiarthritic; antiinfertility; antidiabetic; antiviral; diabetes;
KW   ophthalmological; gene therapy; skin disease; gastrointestinal disorder;
KW   ischaemia; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200104311-A1.
XX
PD   18-JAN-2001.
XX
PF   22-FEB-2000; 2000WO-US04414.
XX
PR   07-JUL-1999;   99US-0143048.
PR   26-JUL-1999;   99US-0145698.
PR   28-JUL-1999;   99US-0146222.
PR   08-SEP-1999;   99WO-US20594.
PR   13-SEP-1999;   99WO-US20944.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   05-OCT-1999;   99WO-US23089.
PR   29-NOV-1999;   99WO-US28214.
PR   30-NOV-1999;   99WO-US28313.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   20-DEC-1999;   99WO-US30999.
PR   05-JAN-2000;   99WO-US00219.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Botstein D,  Desnoyers L,  Eaton DL,  Ferrara N;
PI   Filvaroff E,  Fong S,  Gao W,  Gerber H,  Gerritsen ME,  Goddard A;
PI   Godowski PJ,  Grimaldi CJ,  Gurney AL,  Hillan KJ,  Kljavin IJ;
PI   Mather JP,  Pan J,  Paoni NF,  Roy MA,  Stewart TA,  Tumas D;
PI   Williams PM,  Wood WI;
XX
DR   WPI; 2001-081051/09.
DR   N-PSDB; AAF72423.
XX
PT   Sixty one nucleic acids encoding PRO polypeptides which are useful in
PT   the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung
PT   squamous cell carcinoma) and neurodegenerative diseases (e.g.
PT   Alzheimer's disease) -
XX
PS   Claim 1; Fig 104; 393pp; English.
XX
CC   The present sequence is one of sixty one novel secreted and
CC   transmembrane PRO polypeptides. The PRO polypeptides are
CC   useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung
CC   squamous cell carcinoma), gastrointestinal disorders (e.g.
CC   enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,
CC   Parkinson's disease), wound repair, cardiovascular disorders (e.g.
CC   endometrial bleeding angiogenesis, ischaemias such as coronary
CC   ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,
CC   rheumatoid arthritis, multiple sclerosis), infertility, AIDS and
CC   diabetes and retinal disorders such as retinitis pigmentosum.
CC   The PRO nucleic acids have applications in molecular biology, including
CC   use as hybridization probes, and in chromosome and gene mapping.
XX
SQ   Sequence   640 AA;


  Query Match             100.0%;  Score 2524;  DB 22;  Length 640;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-177;
  Matches  478;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     45 qtcpsvcscsnqfskvicvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhlei 104

Qy     61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    105 lqlsrnhirtieigafnglanlntlelfdnrlttipngafvylsklkelwlrnnpiesip 164

Qy    121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    165 syafnripslrrldlgelkrlsyisegafeglsnlrylnlamcnlreipnltplikldel 224

Qy    181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    225 dlsgnhlsairpgsfqglmhlqklwmiqsqiqviernafdnlqslveinlahnnltllph 284

Qy    241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGEL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    285 dlftplhhlerihlhhnpwncncdilwlswwikdmapsntaccarcntppnlkgryigel 344

Qy    301 DQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    345 dqnyftcyapviveppadlnvtegmaaelkcrastsltsvswitpngtvmthgaykvria 404

Qy    361 VLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    405 vlsdgtlnftnvtvqdtgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmep 464

Qy    421 SQDEARTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEV 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    465 sqdearttdnnvgptpvvdwettnvttsltpqstrstektftipvtdinsgipgidev 522
```

```
RESULT  9
AAB65292
ID   AAB65292 standard; Protein; 640 AA.
XX
AC   AAB65292;
XX
DT   02-APR-2001  (first entry)
XX
DE   Human PRO331 protein sequence SEQ ID NO:501.
XX
KW   Human; secreted and transmembrane protein; PRO; cytostatic;
KW   cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
KW   diagnostic assay.
XX
OS   Homo sapiens.
XX
PN   WO200073454-A1.
XX
PD   07-DEC-2000.
XX
PF   30-MAR-2000; 2000WO-US08439.
XX
PR   02-JUN-1999;   99WO-US12252.
PR   23-JUN-1999;   99US-0141037.
PR   07-JUL-1999;   99US-0143048.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   28-JUL-1999;   99US-0146222.
PR   17-AUG-1999;   99US-0149396.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   08-OCT-1999;   99US-0158663.
PR   30-NOV-1999;   99WO-US28313.
PR   01-DEC-1999;   99WO-US28301.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   05-JAN-2000;  2000WO-US00219.
PR   06-JAN-2000;  2000WO-US00376.
PR   11-FEB-2000;  2000WO-US03565.
PR   18-FEB-2000;  2000WO-US04341.
PR   22-FEB-2000;  2000WO-US04414.
PR   24-FEB-2000;  2000WO-US04914.
PR   24-FEB-2000;  2000WO-US05004.
PR   02-MAR-2000;  2000WO-US05841.
PR   15-MAR-2000;  2000WO-US06884.
PR   20-MAR-2000;  2000WO-US07377.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Botstein D,  Desnoyers L,  Eaton DL;
PI   Ferrara N,  Fong S,  Gerber H,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi CJ,  Gurney AL,  Kljavin IJ,  Napier MA,  Pan J,  Paoni NF;
PI   Roy MA,  Stewart TA,  Tumas D,  Watanabe CK,  Williams PM,  Wood WI;
PI   Zhang Z;
XX
DR   WPI; 2001-032160/04.
DR   N-PSDB; AAF44261.
XX
PT   PRO polynucleotides used to produce polypeptides used to target
PT   bioactive molecules such as toxins, radiolabels or antibodies, to
PT   specific cells, to cause targeted cell death -
XX
PS   Claim 12; Fig 314; 935pp; English.
XX
CC   The present invention describes human secreted and transmembrane PRO
CC   proteins. The PRO proteins have cytostatic activity. The PRO proteins
CC   can be used for targeted delivery of bioactive molecules, such as
CC   toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
CC   sequences, and their fragments, can be used as hybridisation probes, in
CC   chromosomal and gene mapping, and in the generation of anti-sense RNA
CC   and DNA. They may also be used to produce transgenic animals which are
CC   used to develop and screen therapeutically useful reagents. The PRO
CC   nucleotide and protein sequence can be used for tissue typing and in
CC   treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
CC   AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
CC   in the isolation of human PRO sequences. AAF44087 to AAF44269 and
CC   AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC   sequences given in the exemplification of the present invention.
XX
SQ   Sequence   640 AA;


  Query Match             100.0%;  Score 2524;  DB 22;  Length 640;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-177;
  Matches  478;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45 qtcpsvcscsnqfskvicvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhlei 104

Qy      61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     105 lqlsrnhirtieigafnglanlntlelfdnrlttipngafvylsklkelwlrnnpiesip 164

Qy     121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     165 syafnripslrrldlgelkrlsyisegafeglsnlrylnlamcnlreipnltplikldel 224

Qy     181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     225 dlsgnhlsairpgsfqglmhlqklwmiqsqiqviernafdnlqslveinlahnnltllph 284

Qy     241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGEL 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     285 dlftplhhlerihlhhnpwncncdilwlswwikdmapsntaccarcntppnlkgryigel 344

Qy     301 DQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     345 dqnyftcyapviveppadlnvtegmaaelkcrastsltsvswitpngtvmthgaykvria 404

Qy     361 VLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEP 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     405 vlsdgtlnftnvtvqdtgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmep 464

Qy     421 SQDEARTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEV 478
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     465 sqdearttdnnvgptpvvdwettnvttsltpqstrstektftipvtdinsgipgidev 522


RESULT  10
AAB53089
ID   AAB53089 standard; Protein; 640 AA.
XX
AC   AAB53089;
XX
DT   28-FEB-2001  (first entry)
XX
DE   Human angiogenesis-associated protein PRO331, SEQ ID NO:137.
XX
KW   Human; angiogenesis-associated protein; PRO; endothelial cell growth;
KW   cardiac hypertrophy; cardiovascular disorder; endothelial disorder;
KW   angiogenic disorder; atherosclerosis; osteoporosis; hypertension;
KW   myocardial infarction; diabetic retinopathy; rheumatoid arthritis;
KW   Crohn's disease; psoriasis; endometriosis; ulcer; wound healing; cancer;
KW   Alzheimer's disease; Huntington's disease; stroke; drug screening;
KW   gene therapy; transgenic animal.
XX
OS   Homo sapiens.
XX
PN   WO200053753-A2.
XX
PD   14-SEP-2000.
XX
PF   05-JAN-2000; 2000WO-US00219.
XX
PR   08-MAR-1999;   99WO-US05028.
```

```
PR   12-MAR-1999;   99US-0123957.
PR   14-MAY-1999;   99US-0134287.
PR   02-JUN-1999;   99WO-US12252.
PR   23-JUN-1999;   99US-0141037.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   01-SEP-1999;   99WO-US20111.
PR   08-SEP-1999;   99WO-US20594.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   05-OCT-1999;   99WO-US23089.
PR   30-NOV-1999;   99WO-US28313.
PR   30-NOV-1999;   99WO-US28409.
PR   02-DEC-1999;   99WO-US28564.
PR   02-DEC-1999;   99WO-US28565.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Ferrara N,  Gerber H,  Goddard A;
PI   Godowski PJ,  Gurney AL,  Hillan KJ,  Kuo SS,  Mark MR,  Marsters SA;
PI   Paoni NF,  Pitti RM,  Watanabe CK,  Williams PM,  Wood WI;
XX
DR   WPI; 2001-090793/10.
DR   N-PSDB; AAC97475.
XX
PT   New isolated nucleic acid for producing a PRO polypeptide, analyzing
PT   genetic disorders and treating cardiovascular, endothelial or
PT   angiogenic disorders, such as atherosclerosis, wounds or cancer -
XX
PS   Claim 69; Fig 52; 293pp; English.
XX
CC   The invention relates to novel human angiogenesis-associated proteins
CC   designated PRO proteins (AAB53064-B53097), and to nucleic acids encoding
CC   PRO proteins. The invention also relates to vectors and host cells
CC   comprising a PRO nucleic acid, the recombinant production of a PRO
CC   protein, PRO antibodies specific for a PRO protein, fusion proteins
CC   comprising a PRO protein, agonists or antagonists of a PRO protein, and
CC   compounds which inhibit the expression of a PRO gene. The invention
CC   additionally encompasses methods of identifying modulators of PRO
CC   expression or activity; diagnosing a cardiovascular, endothelial or
CC   angiogenic disorder, or a susceptibility to such a disorder by detecting
CC   mutations in a PRO gene, or the expression level of a PRO gene within a
CC   particular tissue; treating a cardiovascular, endothelial or angiogenic
CC   disorder via the administration of a PRO protein, PRO nucleic acid, or
CC   PRO agonist or antagonist; a retroviral gene therapy vector comprising a
CC   PRO nucleic acid; and methods of inhibiting or stimulating endothelial
CC   cell growth, cardiac hypertrophy or PRO-induced angiogenesis via the
CC   administration of a PRO protein, or an agonist or antagonist thereof.
CC   PRO nucleic acids, PRO proteins, antibodies against PRO proteins, PRO
CC   agonists and PRO antagonists may be used as therapeutic agents to treat
CC   cardiovascular, endothelial or angiogenic disorders, such as
CC   atherosclerosis, osteoporosis, myocardial infarction, hypertension,
CC   diabetic retinopathy, rheumatoid arthritis, Crohn's disease, psoriasis,
CC   endometriosis, ulcers, wounds, cancer, Alzheimer's disease, Huntington's
CC   disease, or stroke. PRO nucleic acids are additionally useful in the
CC   recombinant production of PRO proteins, as hybridisation probes to
CC   screen libraries to isolate cDNAs with sequence identity to PRO proteins,
CC   to map genes encoding PRO proteins, to analyse genetic disorders, and in
CC   gene therapy. PRO nucleic acids can also be used to produce transgenic
CC   animals useful for the development and screening of potential
CC   therapeutic agents. The present sequence represents a PRO protein of the
CC   invention.
XX
SQ   Sequence   640 AA;


  Query Match           100.0%;  Score 2524;  DB 22;  Length 640;
  Best Local Similarity  100.0%;  Pred. No. 2.4e-177;
  Matches  478;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     45 qtcpsvcscsnqfskvicvrknlrevpdgistntrllnlhenqiqiikvnsfkhlrhlei 104

Qy     61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    105 lqlsrnhirtieigafnglanlntlelfdnrlttipngafvylsklkelwlrnnpiesip 164

Qy    121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    165 syafnripslrrldlgelkrlsyisegafeglsnlrylnlamcnlreipnltplikldel 224

Qy    181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    225 dlsgnhlsairpgsfqglmhlqklwmiqsqiqviernafdnlqslveinlahnnltllph 284

Qy    241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGEL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    285 dlftplhhlerihlhhnpwncncdilwlswwikdmapsntaccarcntppnlkgryigel 344

Qy    301 DQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    345 dqnyftcyapviveppadlnvtegmaaelkcrastsltsvswitpngtvmthgaykvria 404

Qy    361 VLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    405 vlsdgtlnftnvtvqdtgmytcmvsnsvgnttasatlnvtaatttpfsyfstvtvetmep 464

Qy    421 SQDEARTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEV 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    465 sqdearttdnnvgptpvvdwettnvttsltpqstrstektftipvtdinsgipgidev 522


RESULT  11
AAE13006
ID   AAE13006 standard; Protein; 713 AA.
XX
AC   AAE13006;
XX
DT   28-JAN-2002  (first entry)
XX
DE   Human leucine-rich repeat (LRR) family member protein.
XX
KW   Human; leucine-rich repeat; LRR; 31939 protein; therapy;
KW   cell proliferation; differentiation disorder; cancer; neuronal disorder;
KW   neurological disorder; demyelinating disease; multiple sclerosis;
KW   degenerative disease; Alzheimer's disease; Huntington's disease;
KW   spinocerebellar degeneration; nervous system; bone disorder;
KW   osteoporosis; immune disorder; rheumatoid arthritis; diabetes mellitus;
KW   cardiovascular disorder; liver disorder; viral disease; pain;
KW   metabolic disorder; chromosomal mapping; tissue typing; forensic biology;
KW   cytostatic; nootropic; neuroprotective; anticonvulsant; osteopathic;
KW   antirheumatic; antiarthritic; virucide; analgesic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..38
FT                   /label= Signal_peptide
FT   Protein         39..713
FT                   /note= "Mature human 31939 protein"
FT   Domain          56..85
FT                   /label= N-terminal_LRR_domain
FT   Domain          87..110
FT                   /label= LRR_domain
FT   Domain          111..134
FT                   /label= LRR_domain
FT   Domain          135..158
FT                   /label= LRR_domain
FT   Domain          159..182
FT                   /label= LRR_domain
FT   Domain          183..207
FT                   /label= LRR_domain
FT   Domain          208..229
FT                   /label= LRR_domain
FT   Domain          230..253
```

```
FT                   /label= LRR_domain
FT   Domain          254..277
FT                   /label= LRR_domain
FT   Domain          278..301
FT                   /label= LRR_domain
FT   Domain          311..362
FT                   /label= C-terminal_LRR_domain
FT   Domain          378..438
FT                   /label= Immunoglobulin_domain
XX
PN   WO200175105-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US10380.
XX
PR   31-MAR-2000; 2000US-193919P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Glucksmann MA;
XX
DR   WPI; 2001-662974/76.
DR   N-PSDB; AAD21287.
XX
PT   New polypeptide useful for treating cancer, multiple sclerosis,
PT   Alzheimer's disease, osteoporosis, arthritis, and metabolic and liver
PT   disorders comprises the 31939 polypeptide belonging to the leucine-rich
PT   repeat family -
XX
PS   Claim 9; Fig 1A-1C; 117pp; English.
XX
CC   The present sequence is a human leucine-rich repeat (LRR) family
CC   member (referred as 31939) protein. The 31939 protein is useful for
CC   treating cell proliferative or differentiative disorder, e.g., cancer
CC   disorders or neuronal disorders, neurological disorders, demyelinating
CC   diseases such as multiple sclerosis, degenerative diseases such as
CC   Alzheimer's disease and Huntington's disease, spinocerebellar
CC   degenerations, disorders of the central or peripheral nervous system,
CC   bone disorders such as osteoporosis, immune disorders such as
CC   rheumatoid arthritis, diabetes mellitus, cardiovascular disorders,
CC   liver disorders, viral diseases, pain or metabolic disorders. They
CC   are also useful as markers of disorders or disease states, or for
CC   precursors or markers for the predisposition of disease states. They
CC   are useful in screening and detection assays such as chromosomal
CC   mapping, tissue typing, and forensic biology.
XX
SQ   Sequence   713 AA;
```

```
  Query Match             71.1%;  Score 1793.5;  DB 22;  Length 713;
  Best Local Similarity   64.9%;  Pred. No. 1.5e-123;
  Matches  335;  Conservative   68;  Mismatches   74;  Indels   39;  Gaps    6;

Qy       2 TCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEIL 61
           :||  |||||| |:||| |::| |||  |  ||| ||| || ||:|: ::||||||||||
Db      56 scpvacscsnqasrvictrrdlaevpasipvntrylnlqengiqvirtdtfkhlrhleil 115

Qy      62 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 121
           |||:| :| ||:|||||| :|||||||||||||:|  || |||||:|||||||||||||
Db     116 qlsknlvrkievgafnglpslntlelfdnrlttvptqafeylsklrelwlrnnpiesips 175

Qy     122 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELD 181
           |||||:||||||||||||| |||| ||||| ||||||| ||||::||||| |::|:||:
Db     176 yafnrvpslrrldlgelkrleyiseaafeglvnlrylnlgmcnlkdipnltalvrleele 235

Qy     182 LSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHD 241
           |||| |  |||||||||  ||:|||:: :|:  |||||||:|:|| |:||:|||  ||||
Db     236 lsgnrldlirpgsfqgltslrklwlmhaqvatiernafddlksleelnlshnnlmslphd 295

Qy     242 LFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELD 301
           |||||| |||:||:||||:||||:|||||:|:  |||| |||||: |  ||||||||||
Db     296 lftplhrlervhlnhnpwhcncdvlwlswwlketvpsnttccarchapaglkgryigeld 355

Qy     302 QNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIAV 361
           |::||||||||||| ||||||||||||||  ||:|||:|:|||||:||||:|:|||:|
Db     356 qshftcyapvivepptdlnvtegmaaelkcrtgtsmtsvnwltpngtlmthgsyrvrisv 415

Qy     362 LSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTP-------------- 406
           | ||||||||||||||| |||||:|| ||||||||||:|
Db     416 lhdgtlnftnvtvqdtgqytcmvtnsagnttasatlnvsavdpvaaggtgsgggpggsg 475

Qy     407 --------FSYFSTVTVETME--PSQD--EARTTDNN-VGPTPVVDW----------ETT 443
                   ::||:|||||||:|  | ::  : | |:    |||    |           :
Db     476 gvgggsggytyfttvtvetletqpgeealqprgtekeppgpttdgvwgggrpgdaagpas 535

Qy     444 NVTTSLTPQSTRSTEKTFTIPVTDI-NSGIPGIDEV 478
           : ||:  |:|:| ||| ||:|:||:  : :  :|:|
Db     536 ssttapaprssrptekaftvpitdvtenalkdlddv 571


RESULT  12
AAY28806
ID   AAY28806 standard; Protein; 653 AA.
XX
AC   AAY28806;
XX
DT   17-JAN-2000  (first entry)
XX
DE   cc359_4 secreted protein.
XX
KW   clone cc359_4; cc359_4 protein; human adult brain cDNA library;
KW   secreted protein; signal peptide; transmembrane domain; leucine zipper;
KW   hydrophobic nature; TopPredII computer program; cytokine; tissue growth;
KW   nutritional activity; cell proliferation; immune stimulation;
KW   chaoptin domain; immune suppression; hematopoiesis regulation;
KW   tumour inhibition.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         29..41
FT                   /note= "Predicted signal peptide"
FT   Protein         42..653
FT                   /note= "Mature cc359_4 protein"
FT   Domain          94..115
FT                   /label= Chaoptin_domain-1
FT                   /note= "Composed of 41 potentially amphipathic repeats"
FT   Domain          118..139
FT                   /label= Chaoptin_domain-2
FT                   /note= "Composed of 41 potentially amphipathic repeats"
FT   Domain          142..163
FT                   /label= Chaoptin_domain-3
FT                   /note= "Composed of 41 potentially amphipathic repeats"
FT   Domain          261..282
FT                   /label= Chaoptin_domain-4
FT                   /note= "Composed of 41 potentially amphipathic repeats"
XX
PN   WO9950405-A1.
XX
PD   07-OCT-1999.
XX
PF   30-MAR-1999;   99WO-US06946.
XX
PR   31-MAR-1998;   98US-0080110.
PR   29-MAR-1999;   99US-0280591.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  Mccoy JM,  LaVallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M,  Agostino MJ,  Steininger RJ;
XX
DR   WPI; 1999-610849/52.
DR   N-PSDB; AAX90848.
XX
PT   Polynucleotides encoding secreted human proteins, derived from human
```

```
PT   adult brain, human fetal brain, human fetal kidney, and human adult
PT   blood cDNA libraries -
XX
PS   Claim 1; Page 97-99; 122pp; English.
XX
CC   The present sequence is the cc359_4 secreted protein encoded by the cDNA
CC   clone cc359_4. cc359_4 was isolated from a human adult brain cDNA library
CC   using methods specific for secreted protein cDNAs. The leader sequence or
CC   signal peptide acts as a transmembrane domain due to its hydrophobic
CC   nature. The TopPredII computer program predicts five potential
CC   transmembrane domains centered around amino acids 20, 410, 490, 530 and
CC   590. This protein has a leucine zipper motif. The polynucleotide and
CC   protein may effect nutritional activity, cytokine and cell proliferation,
CC   immune stimulation or suppression, hematopoiesis regulation, tissue
CC   growth, tumour inhibition etc.
XX
SQ   Sequence   653 AA;


  Query Match           62.7%;  Score 1581.5;  DB 20;  Length 653;
  Best Local Similarity   60.9%;  Pred. No. 5.1e-108;
  Matches  298;  Conservative   77;  Mismatches   93;  Indels   21;  Gaps    8;

Qy       1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
           | ||||||||||||||:| |: | ||| || :||| ||| || ||:|: ::|:|| |||:
Db      44 qncpsvcscsnqfskvvctrrglsevpqgipsntrylnlmenniqmiqadtfrhlhhlev 103

Qy      61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
           ||| || || ||:||||||||:|||||||| || ||:||| ||||||:|||||||||||||
Db     104 lqlgrnsirqievgafnglaslntlelfdnwltvipsgafeylsklrelwlrnnpiesip 163

Qy     121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
           ||||||:||| ||||||||:| |||||||||| ||:|||| |||::::|||||||: |:||
Db     164 syafnrvpslmrldlgelkkleyisegafeglfnlkylnlgmcnikdmpnltplvgleel 223

Qy     181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
           ::||||   |||||| ||  |:|||:: ||: :|||||||  | ||||:|||||||: |||
Db     224 emsgnhfpeirpgsfhglsslkklwvmnsqvsliernafdglaslvelnlahnnlsslph 283

Qy     241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGEL 300
           |||||| :|  :|||||||||:||||||:||:::  |:|: || ||: | :::|||:*|:
Db     284 dlftplrylvelhlhhnpwncdcdilwlawwlreyiptnstccgrchapmhmrgrylvev 343

Qy     301 DQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIA 360
           ||  | | || |:: | |||::||  |||||| :  ::|| |: |||||::| :   ||:
Db     344 dqasfqcsapfimdaprdlnisegrmaelkcr-tppmssvkwllpngtvlshasrhpris 402

Qy     361 VLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAA--TTTPFSYFSTVTVETM 418
           ||:||||||::| : |||:|||||:|  ||: ||| |||: |   |: :|:|:|||||
Db     403 vlndgtlnfshvllsdtgvytcmvtnvagnsnasaylnvstaelntsnysfftttvtvett 462

Qy     419 EPSQDEARTTDNNVGPTPVVDWETTNV------TTSLTP--QSTRSTEKTFTIPVTDINS 470
           | | ::   |    | |     ||:       ||| |   |:||   |   :| ||
Db     463 eisped---ttrkykpvp-----ttstgyqpayttsttvliqttr-vpkqvavpatdttd 513

Qy     471 GI-PGIDEV 478
             :   :|||
Db     514 kmqtsldev 522


RESULT  13
AAB24073
ID   AAB24073 standard; Protein; 653 AA.
XX
AC   AAB24073;
XX
DT   29-JAN-2001  (first entry)
XX
DE   Human PRO1111 protein sequence SEQ ID NO:46.
XX
KW   Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;
KW   proliferation; tumourigenesis; identification; cancer; cytostatic;
KW   nootropic; neuroprotective; antiinflammatory; immunosuppressive;
KW   immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;
KW   neuronal disorder; glial disorder; astrocytal disorder; angiogenic;
KW   hypothalamic disorder; glandular disorder; macrophagal disorder;
KW   epithelial disorder; stromal disorder; blastocoelic disorder;
KW   inflammatory disorder; immunologic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200053755-A2.
XX
PD   14-SEP-2000.
XX
PF   06-JAN-2000; 2000WO-US00376.
XX
PR   08-MAR-1999;   99WO-US05028.
PR   02-JUN-1999;   99WO-US12252.
PR   23-JUN-1999;   99US-0141037.
PR   07-JUL-1999;   99US-0143048.
PR   26-JUL-1999;   99US-0145698.
PR   30-NOV-1999;   99WO-US28313.
PR   20-DEC-1999;   99WO-US30911.
PR   05-JAN-2000; 2000WO-US00219.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Goddard A,  Gurney AL,  Hillan KJ,  Roy MA;
PI   Watanabe CK,  Wood WI;
XX
DR   WPI; 2000-572270/53.
DR   N-PSDB; AAC58383.
XX
PT   Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
PT   treatment, diagnosis and prevention of cancer -
XX
PS   Claim 61; Fig 34; 286pp; English.
XX
CC   The present invention describes an isolated antibody that binds to
CC   one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535,
CC   PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,
CC   PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,
CC   PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,
CC   PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
CC   growth. The PRO polypeptides and nucleotides are useful in the
CC   treatment, diagnosis and prevention of cancer. The antibodies and other
CC   anti-tumour compounds maybe used to treat various conditions, including
CC   those characterised by overexpression and/or activation of the amplified
CC   PRO genes. Exemplary conditions or disorders to be treated with such
CC   antibodies and other compounds include benign or malignant tumours
CC   (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
CC   colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
CC   carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
CC   leukaemias and lymphoid malignancies, other disorders such as neuronal,
CC   glial, astrocytal, hypothalamic and other glandular, macrophagal,
CC   epithelial, stromal and blastocoelic disorders, and inflammatory,
CC   angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
CC   primers and hybridisation probes used in the isolation of the human PRO
CC   sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human
CC   PRO polynucleotide and protein sequences given in the exemplification of
CC   the present invention.
XX
SQ   Sequence   653 AA;


  Query Match           62.7%;  Score 1581.5;  DB 21;  Length 653;
  Best Local Similarity   60.9%;  Pred. No. 5.1e-108;
  Matches  298;  Conservative   77;  Mismatches   93;  Indels   21;  Gaps    8;

Qy       1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
           | ||||||||||||||:| |: | ||| || :||| ||| || ||:|: ::|:|| |||:
Db      44 qncpsvcscsnqfskvvctrrglsevpqgipsntrylnlmenniqmiqadtfrhlhhlev 103

Qy      61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
           ||| || || ||:||||||||:|||||||| || ||:||| ||||||:|||||||||||||
```

```
Db    104 lqlgrnsirqievgafnglaslntlelfdnwltvipsgafeylsklrelwlrnnpiesip 163

Qy    121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180

Db    164 syafnrvpslmrldlgelkkleyisegafeglfnlkylnlgmcnikdmpnltplvgleel 223

Qy    181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240

Db    224 emsgnhfpeirpgsfhglsslkklwvmnsqvsliernafdglaslvelnlahnnlsslph 283

Qy    241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGEL 300

Db    284 dlftplrylvelhlhhnpwncdcdilwlawwlreyiptnstccgrchapmhmrgrylvev 343

Qy    301 DQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIA 360

Db    344 dqasfqcsapfimdaprdlnisegrmaelkcr-tppmssvkwllpngtvlshasrhpris 402

Qy    361 VLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAA--TTTPFSYFSTVTVETM 418

Db    403 vlndgtlnfshvllsdtgvytcmvtnvagnsnasaylnvstaelntsnysffttvtvett 462

Qy    419 EPSQDEARTTDNNVGPTPVVDWETTNV------TTSLTP--QSTRSTEKTFTIPVTDINS 470

Db    463 eisped---ttrkykpvp-----ttstgyqpayttsttvliqttr-vpkqvavpatdttd 513

Qy    471 GI-PGIDEV 478

Db    514 kmqtsldev 522


RESULT 14
AAY66694
ID   AAY66694 standard; protein; 653 AA.
XX
AC   AAY66694;
XX
DT   05-APR-2000  (first entry)
XX
DE   Membrane-bound protein PRO1111.
XX
KW   Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW   pharmaceutical; receptor immunoadhesin; gene mapping.
XX
OS   Homo sapiens.
XX
PN   WO9963088-A2.
XX
PD   09-DEC-1999.
XX
PF   02-JUN-1999;   99WO-US12252.
XX
PR   02-JUN-1998;   98US-0087607.
PR   02-JUN-1998;   98US-0087609.
PR   02-JUN-1998;   98US-0087759.
PR   03-JUN-1998;   98US-0087827.
PR   04-JUN-1998;   98US-0088021.
PR   04-JUN-1998;   98US-0088025.
PR   04-JUN-1998;   98US-0088028.
PR   04-JUN-1998;   98US-0088029.
PR   04-JUN-1998;   98US-0088030.
PR   04-JUN-1998;   98US-0088033.
PR   04-JUN-1998;   98US-0088326.
PR   05-JUN-1998;   98US-0088167.
PR   05-JUN-1998;   98US-0088202.
PR   05-JUN-1998;   98US-0088212.
PR   05-JUN-1998;   98US-0088217.
PR   09-JUN-1998;   98US-0088655.
PR   10-JUN-1998;   98US-0088722.
PR   10-JUN-1998;   98US-0088730.
PR   10-JUN-1998;   98US-0088734.
PR   10-JUN-1998;   98US-0088738.
PR   10-JUN-1998;   98US-0088740.
PR   10-JUN-1998;   98US-0088741.
PR   10-JUN-1998;   98US-0088742.
PR   10-JUN-1998;   98US-0088810.
PR   10-JUN-1998;   98US-0088811.
PR   10-JUN-1998;   98US-0088824.
PR   10-JUN-1998;   98US-0088825.
PR   10-JUN-1998;   98US-0088826.
PR   11-JUN-1998;   98US-0088858.
PR   11-JUN-1998;   98US-0088861.
PR   11-JUN-1998;   98US-0088863.
PR   11-JUN-1998;   98US-0088876.
PR   12-JUN-1998;   98US-0089090.
PR   12-JUN-1998;   98US-0089105.
PR   16-JUN-1998;   98US-0089440.
PR   16-JUN-1998;   98US-0089512.
PR   16-JUN-1998;   98US-0089514.
PR   17-JUN-1998;   98US-0089532.
PR   17-JUN-1998;   98US-0089538.
PR   17-JUN-1998;   98US-0089598.
PR   17-JUN-1998;   98US-0089599.
PR   17-JUN-1998;   98US-0089600.
PR   17-JUN-1998;   98US-0089653.
PR   18-JUN-1998;   98US-0089801.
PR   18-JUN-1998;   98US-0089907.
PR   18-JUN-1998;   98US-0089908.
PR   19-JUN-1998;   98US-0089947.
PR   19-JUN-1998;   98US-0089948.
PR   19-JUN-1998;   98US-0089952.
PR   22-JUN-1998;   98US-0090246.
PR   22-JUN-1998;   98US-0090252.
PR   22-JUN-1998;   98US-0090254.
PR   23-JUN-1998;   98US-0090349.
PR   23-JUN-1998;   98US-0090355.
PR   24-JUN-1998;   98US-0090429.
PR   24-JUN-1998;   98US-0090431.
PR   24-JUN-1998;   98US-0090435.
PR   24-JUN-1998;   98US-0090444.
PR   24-JUN-1998;   98US-0090445.
PR   24-JUN-1998;   98US-0090461.
PR   24-JUN-1998;   98US-0090472.
PR   24-JUN-1998;   98US-0090535.
PR   24-JUN-1998;   98US-0090538.
PR   24-JUN-1998;   98US-0090540.
PR   24-JUN-1998;   98US-0090557.
PR   25-JUN-1998;   98US-0090676.
PR   25-JUN-1998;   98US-0090678.
PR   25-JUN-1998;   98US-0090688.
PR   25-JUN-1998;   98US-0090690.
PR   25-JUN-1998;   98US-0090691.
PR   25-JUN-1998;   98US-0090694.
PR   25-JUN-1998;   98US-0090695.
PR   25-JUN-1998;   98US-0090696.
PR   26-JUN-1998;   98US-0090862.
PR   26-JUN-1998;   98US-0090863.
PR   01-JUL-1998;   98US-0091358.
PR   01-JUL-1998;   98US-0091360.
PR   01-JUL-1998;   98US-0091544.
PR   02-JUL-1998;   98US-0091478.
PR   02-JUL-1998;   98US-0091486.
PR   02-JUL-1998;   98US-0091519.
PR   02-JUL-1998;   98US-0091626.
PR   02-JUL-1998;   98US-0091628.
PR   02-JUL-1998;   98US-0091633.
PR   02-JUL-1998;   98US-0091646.
PR   02-JUL-1998;   98US-0091673.
PR   07-JUL-1998;   98US-0091978.
PR   07-JUL-1998;   98US-0091982.
PR   09-JUL-1998;   98US-0092182.
PR   10-JUL-1998;   98US-0092472.
PR   20-JUL-1998;   98US-0093339.
PR   30-JUL-1998;   98US-0094651.
PR   04-AUG-1998;   98US-0095282.
PR   04-AUG-1998;   98US-0095285.
```

```
PR   04-AUG-1998;   98US-0095301.
PR   04-AUG-1998;   98US-0095302.
PR   04-AUG-1998;   98US-0095318.
PR   04-AUG-1998;   98US-0095321.
PR   04-AUG-1998;   98US-0095325.
PR   10-AUG-1998;   98US-0095916.
PR   10-AUG-1998;   98US-0095929.
PR   10-AUG-1998;   98US-0096012.
PR   11-AUG-1998;   98US-0096143.
PR   11-AUG-1998;   98US-0096146.
PR   12-AUG-1998;   98US-0096329.
PR   17-AUG-1998;   98US-0096757.
PR   17-AUG-1998;   98US-0096766.
PR   17-AUG-1998;   98US-0096768.
PR   17-AUG-1998;   98US-0096773.
PR   17-AUG-1998;   98US-0096791.
PR   17-AUG-1998;   98US-0096867.
PR   17-AUG-1998;   98US-0096891.
PR   17-AUG-1998;   98US-0096894.
PR   17-AUG-1998;   98US-0096895.
PR   17-AUG-1998;   98US-0096897.
PR   18-AUG-1998;   98US-0096949.
PR   18-AUG-1998;   98US-0096950.
PR   18-AUG-1998;   98US-0096959.
PR   18-AUG-1998;   98US-0096960.
PR   18-AUG-1998;   98US-0097022.
PR   19-AUG-1998;   98US-0097141.
PR   20-AUG-1998;   98US-0097218.
PR   24-AUG-1998;   98US-0097661.
PR   26-AUG-1998;   98US-0097951.
PR   26-AUG-1998;   98US-0097952.
PR   26-AUG-1998;   98US-0097954.
PR   26-AUG-1998;   98US-0097955.
PR   26-AUG-1998;   98US-0097971.
PR   26-AUG-1998;   98US-0097974.
PR   26-AUG-1998;   98US-0097978.
PR   26-AUG-1998;   98US-0097979.
PR   26-AUG-1998;   98US-0097986.
PR   26-AUG-1998;   98US-0098014.
PR   31-AUG-1998;   98US-0098525.
PR   16-SEP-1998;   98US-0100634.
PR   12-JAN-1999;   99US-0115565.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker K,  Chen J,  Goddard A,  Gurney AL,  Smith V,  Watanabe CK;
PI   Wood WI,  Yuan J;
XX
DR   WPI; 2000-072883/06.
DR   N-PSDB; AAZ65033.
XX
PT   Membrane-bound proteins and related nucleotide sequences  -
XX
PS   claim 12; Fig 157; 822pp; English.
XX
CC   The invention provides membrane-bound PRO polypeptides and
CC   polynucleotides encoding them. The PRO sequences of the invention were
CC   identified based on extracellular domain homology screening. The PRO
CC   sequences have homology with proteins including LDL receptors, TIE
CC   ligands and various enzymes. The membrane-bound proteins and receptor
CC   molecules are useful as pharmaceutical and diagnostic agents. Receptor
CC   immunoadhesins, for instance, can be used as therapeutic agents to block
CC   receptor-ligand interactions. The membrane-bound proteins can also be
CC   employed for screening of potential peptide or small molecule inhibitors
CC   of the relevant receptor/ligand interaction. The PRO encoding sequences
CC   are useful as hybridization probes, in chromosome and gene mapping and in
CC   the generation of antisense RNA and DNA. PRO nucleic acid sequences
CC   will also be useful for the preparation of PRO polypeptides, especially
CC   by recombinant techniques.
XX
SQ   Sequence   653 AA;
```

```
  Query Match           62.7%;  Score 1581.5;  DB 21;  Length 653;
  Best Local Similarity   60.9%;  Pred. No. 5.1e-108;
  Matches  298;  Conservative   77;  Mismatches   93;  Indels   21;  Gaps    8;

Qy       1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
           | ||||||||||||||:| |: | ||| || :||| ||| || ||:|: ::|:|| |||:
Db      44 qncpsvcscsnqfskvvctrrglsevpqgipsntrylnlmenniqmiqadtfrhlhhlev 103

Qy      61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
           ||| || || ||:|||||||:||||||||| || ||:||| |||||:||||||||||||||
Db     104 lqlgrnsirqievgafnglaslntlelfdnwltvipsgafeylsklrelwlrnnpiesip 163

Qy     121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
           ||||||:||| |||||||:| |||||||||| ||:|||| |||:::::|||||: |:||
Db     164 syafnrvpslmrldlgelkkleyisegafeglfnlkylnlgmcnikdmpnltplvgleel 223

Qy     181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
           ::||||   |||||| ||  |:|||:: ||: :||||||| | ||||:|||||||: |||
Db     224 emsgnhfpeirpgsfhglsslkklwvmnsqvsliernafdglaslvelnlahnnlsslph 283

Qy     241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGEL 300
           |||||| :|  :|||||||||:||||||:||:::  |:|: || ||: | :::|||: |:
Db     284 dlftplrylvelhlhhnpwncdcdilwlawwlreyiptnstccgrchapmhmrgrylvev 343

Qy     301 DQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIA 360
           ||  | | || |:: | |||::||  |||||| :  ::|| |: |||||::| :   ||:
Db     344 dqasfqcsapfimdaprdlnisegrmaelkcr-tppmssvkwllpngtvlshasrhpris 402

Qy     361 VLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAA--TTTPFSYFSTVTVETM 418
           ||:||||||::| : |||:|||||:|  ||: ||| |||: |   |: :|:|:|||||
Db     403 vlndgtlnfshvllsdtgvytcmvtnvagnsnasaylnvstaelntsnysffttvtvett 462

Qy     419 EPSQDEARTTDNNVGPTPVVDWETTNV------TTSLTP--QSTRSTEKTFTIPVTDINS 470
           | | ::   |     | |     ||:       ||| |   |:||   |   :| ||
Db     463 eisped---ttrkykpvp-----ttstgyqpayttsttvliqttr-vpkqvavpatdttd 513

Qy     471 GI-PGIDEV 478
             :   :|||
Db     514 kmqtsldev 522


RESULT  15
AAE09438
ID   AAE09438 standard; Protein; 653 AA.
XX
AC   AAE09438;
XX
DT   19-NOV-2001  (first entry)
XX
DE   Human sbgPRO331a protein.
XX
KW   Human; Alzheimer's disease; amyotrophic lateral sclerosis;
KW   ALS; Zollinger-Ellison syndrome; immune system disease; schizophrenia;
KW   inflammation; haematopoietic disease; anxiety; feeding disorder; aging;
KW   anorexia; depression; cardiovascular disease; sleep disorder; seizure;
KW   memory alteration; migraine; stroke; asthma; neuropathy; hypoglycaemia;
KW   sexual disorder; growth abnormality; infection; autoimmune disease;
KW   rheumatoid arthritis; cataractogenesis; angiogenesis; atherosclerosis;
KW   cerebral ischaemia; cirrhosis; Huntington's disease; Hodgson's disease;
KW   hypercholesterolaemia; headache; amnesia; cardiac arrhythmia; obesity;
KW   diabetes mellitus; glomerulonephritis; renovascular hypertension;
KW   cancer; vaccine; gene therapy; sbgPRO331a gene.
XX
OS   Homo sapiens.
XX
PN   WO200160850-A1.
XX
PD   23-AUG-2001.
XX
PF   14-FEB-2001; 2001WO-US04703.
XX
PR   14-FEB-2000; 2000US-0182172.
PR   29-FEB-2000; 2000US-0186084.
```

PR 18-APR-2000; 2000US-0198583.
PR 04-OCT-2000; 2000US-0237963.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
XX
PI Agarwal P, Kabnick KS, Murdoch PR, Rizvi SK, Smith RF, Xiang Z;
XX
DR WPI; 2001-536566/59.
DR N-PSDB; AAD16345.
XX
PT New secreted and membrane associated polypeptides for treating
PT Alzheimer's disease, psoriasis, cancer, enterocolitis, sleep and sexual
PT disorders, stroke, and asthma -
XX
PS Claim 1; Page 56-58; 94pp; English.
XX
CC The present sequence is a human sbgPRO331a protein,
CC a membrane bound protein of the invention.
CC The invention relates to secreted and membrane associated polypeptides
CC and nucleic acid molecules encoding such polypeptides. Sequences of the
CC invention are useful for treating diseases such as Alzheimer's disease,
CC amyotrophic lateral sclerosis (ALS), Zollinger-Ellison syndrome, diseases
CC of the immune system, haematopoietic disease, inflammation, anxiety,
CC schizophrenia, feeding disorders, anorexia, depression, social, sexual
CC and rewarded behaviour, cardiovascular disease, sleep disorder, learning
CC and memory alteration and altered immune response, seizure, migraine,
CC cancer, stroke, asthma, neuropathy, aging, sexual disorders, treatment
CC of transsexuals, growth abnormalities, obesity, infections, autoimmune
CC diseases (e.g. rheumatoid arthritis), cataractogenesis, angiogenesis,
CC disorders associated with healthy maintenance of gastric mucosa and
CC repair of acute and chronic mucosal lesion, lung carcinoma, cerebral
CC ischaemia, atherosclerosis, cirrhosis, Huntington's disease, headache,
CC amnesia, multiple sclerosis, Hodgson's disease, dilated cardiomyopathy,
CC congestive heart failure, cardiac arrhythmias, hypercholesterolaemia,
CC viral and non-viral hepatitis, type I and type II diabetes mellitus,
CC glomerulonephritis, renovascular hypertension, hypoglycaemia, periodic
CC paralyses, tendinitis and malignant hyperthermia. Polypeptides of the
CC invention are used to identify membrane bound and soluble receptors.
CC They are also useful as vaccines for inducing an immunological response
CC in a mammal. Polynucleotides of the invention are used in gene therapy.
CC They are also valuable for chromosome localisation studies and tissue
CC expression studies.
XX
SQ Sequence 653 AA;

Query Match 62.7%; Score 1581.5; DB 22; Length 653;
Best Local Similarity 60.9%; Pred. No. 5.1e-108;
Matches 298; Conservative 77; Mismatches 93; Indels 21; Gaps 8;

```
Qy       1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
           | ||||||||||||||:| |: | ||| || :||| ||| || ||:|: ::|:|| |||:
Db      44 qncpsvcscsnqfskvvctrrglsevpqgipsntrylnlmenniqmiqadtfrhlhhlev 103

Qy      61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
           ||| || || ||:||||||:|||||||| || ||:||| |||||:|||||||||||||
Db     104 lqlgrnsirqievgafnglaslntlelfdnwltvipsgafeylsklrelwlrnnpiesip 163

Qy     121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
           ||||||:||| |||||||:| |||||||||| ||:|||| |||::::||||||: |:||
Db     164 syafnrvpslmrldlgelkkleyisegafeglfnlkylnlgmcnikdmpnltplvgleel 223

Qy     181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
           ::|||   |||||| ||  |:|||:: ||: :|||||||| | ||||:|||||||: |||
Db     224 emsgnhfpeirpgsfhglsslkklwvmnsqvsliernafdglaslvelnlahnnlsslph 283

Qy     241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGEL 300
           |||||| :|  :|||||||||:||||||:||:::  |:|: || ||: | :::|||: |:
Db     284 dlftplrylvelhlhhnpwncdcdilwlawwlreyiptnstccgrchapmhmrgrylvev 343

Qy     301 DQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIA 360
           ||  | | || |:: | |||::||  |||||| :  ::|| |: ||||||:| :   ||:
```

```
Db     344 dqasfqcsapfimdaprdlnisegrmaelkcr-tppmssvkwllpngtvlshasrhpris 402

Qy     361 VLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAA--TTTPFSYFSTVTVETM 418
           ||:||||||::| : |||:|||||:|  ||: ||| |||: |   |: :|:|:||||||
Db     403 vlndgtlnfshvllsdtgvytcmvtnvagnsnasaylnvstaelntsnysfpttvtvett 462

Qy     419 EPSQDEARTTDNNVGPTPVVDWETTNV------TTSLTP--QSTRSTEKTFTIPVTDINS 470
           | | ::  |     | |     ||:       ||| |   |:||   |   :| ||
Db     463 eisped---ttrkykpvp-----ttstgyqpayttsttvliqttr-vpkqvavpatdttd 513

Qy     471 GI-PGIDEV 478
            :   :|||
Db     514 kmqtsldev 522
```

Search completed: August 26, 2002, 15:46:13
Job time: 260 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 2002, 15:44:38 ; Search time 20.91 Seconds
(without alignments)
2196.590 Million cell updates/sec

Title: US-09-905-056-292_COPY_45_522
Perfect score: 2524
Sequence: 1 QTCPSVCSCSNQFSKVICVR.....KTFTIPVTDINSGIPGIDEV 478

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1151.5	45.6	421	2	T46266	hypothetical prote
2	402.5	15.9	1091	2	A58532	glial cell membran
3	396	15.7	707	2	JC7763	neuronal leucine-r
4	394	15.6	361	2	A53860	chondroadherin pre
5	389.5	15.4	1531	2	T42218	slit-1 protein hom
6	365.5	14.5	1523	2	T13953	MEGF5 protein - ra
7	350	13.9	1469	2	B36665	slit protein 2 pre
8	350	13.9	1480	2	A36665	slit protein 1 pre
9	345	13.7	789	2	T28714	hypothetical prote
10	345	13.7	1355	2	T28715	hypothetical prote
11	344.5	13.6	1025	2	T42626	secreted leucine-r
12	329.5	13.1	560	2	A60164	platelet membrane
13	329.5	13.1	605	2	JC5239	insulin-like growt
14	314.5	12.5	605	2	A41915	insulin-like growt
15	313	12.4	1535	2	S46224	peroxidasin - frui
16	309	12.2	603	2	JC1282	insulin-like growt
17	302.5	12.0	603	2	JC6128	insulin-like growt
18	302	12.0	359	1	NBHUC8	decorin precursor
19	293	11.6	360	2	S06280	decorin precursor
20	291	11.5	357	2	S24317	decorin precursor
21	290.5	11.5	594	2	T23841	hypothetical prote
22	289	11.5	360	2	I47020	decorin - rabbit
23	286.5	11.4	536	2	A34901	lysine carboxypept
24	286	11.3	354	2	A55454	decorin precursor
25	283.5	11.2	420	2	A53531	oncofetal trophobl
26	283	11.2	354	2	S29145	decorin precursor
27	282	11.2	626	1	NBHUIA	platelet glycoprot
28	281.5	11.2	610	2	T23836	hypothetical prote
29	281.5	11.2	1389	2	T13852	gene wheeler prote
30	278.5	11.0	907	2	JE0176	orphan G protein-c
31	276.5	11.0	1385	2	T13887	tlr protein - frui
32	273.5	10.8	907	2	JG0193	G protein-coupled
33	271.5	10.8	368	1	BGHUN	biglycan precursor
34	270.5	10.7	369	2	S20811	proteoglycan I - m
35	270.5	10.7	369	2	S32793	biglycan precursor
36	269.5	10.7	369	2	S32559	biglycan precursor
37	265.5	10.5	653	2	T25194	hypothetical prote
38	264.5	10.5	458	2	T19941	hypothetical prote
39	264	10.5	603	2	T24315	hypothetical prote
40	262	10.4	680	2	T19939	hypothetical prote
41	258.5	10.2	562	2	T34319	hypothetical prote
42	253	10.0	1066	2	T15864	hypothetical prote
43	252	10.0	575	2	T29972	hypothetical prote
44	250.5	9.9	382	2	I39068	proline- arginine-
45	246.5	9.8	1134	1	A29944	chaoptin precursor

ALIGNMENTS

RESULT 1
T46266
hypothetical protein DKFZp761A179.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T46266
R;Blum, H.; Bauersachs, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23034
A;Accession: T46266
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-421 <AAA>
A;Cross-references: EMBL:AL137451
A;Experimental source: adult amygdala; clone DKFZp761A179
C;Genetics:
A;Note: DKFZp761A179.1

Query Match 45.6%; Score 1151.5; DB 2; Length 421;
Best Local Similarity 59.8%; Pred. No. 3.3e-72;
Matches 213; Conservative 53; Mismatches 51; Indels 39; Gaps 6;

```
Qy    162 MCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDN 221
          ||||::|||||  |::|:||:|||| |  |||||||||  |:|||:: :|:  |||||||:
Db      1 MCNLKDIPNLTALVRLEELELSGNRLDLIRPGSFQGLTSLRKLWLMHAQVATIERNAFDD 60

Qy    222 LQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTA 281
          |:|| |:||:|||| ||||||||| |||:||:||||:||||:||||||:|:  ||||
Db     61 LKSLEELNLSHNNLMSLPHDLFTPLHRLERVHLNHNPWHCNCDVLWLSWWLKETVPSNTT 120

Qy    282 CCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVS 341
          |||||: |  |||||||||||::||||||||||||| ||||||||||||||  ||:|||:
Db    121 CCARCHAPAGLKGRYIGELDQSHFTCYAPVIVEPPTDLNVTEGMAAELKCRTGTSMTSVN 180

Qy    342 WITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTA 401
          |:||||:||||:|:|||:|| ||||||||||||||| |||||:|| ||||||||||:|
Db    181 WLTPNGTLMTHGSYRVRISVLHDGTLNFTNVTVQDTGQYTCMVTNSAGNTTASATLNVSA 240

Qy    402 ATTTP----------------------FSYFSTVTVETME--PSQD--EARTTDNN-VG 433
                                       ::||:||||||:|  | ::  : | |:    |
Db    241 VDPVAAGGTGSGGGGPGGSGGVGGGSGGYTYFTTVTVETLETQPGEEALQPRGTEKEPPG 300

Qy    434 PTPVVDW----------ETTNVTTSLTPQSTRSTEKTFTIPVTDI-NSGIPGIDEV 478
          ||    |            :: ||:  |:|:| ||| ||:|:||:  : :  :|:|
Db    301 PTTDGVWGGGRPGDAAGPASSSTTAPAPRSSRPTEKAFTVPITDVTENALKDLDDV 356
```

RESULT 2
A58532
glial cell membrane glycoprotein LIG-1 precursor - mouse
C;Species: Mus musculus (house mouse)

```
C;Date: 11-Apr-1997 #sequence_revision 11-Apr-1997 #text_change 05-Nov-1999
C;Accession: A58532
R;Suzuki, Y.; Sato, N.; Tohyama, M.; Wanaka, A.; Takagi, T.
J. Biol. Chem. 271, 22522-22527, 1996
A;Title: cDNA cloning of a novel membrane glycoprotein that is expressed specifically in
A;Reference number: A58532; MUID:96394313
A;Accession: A58532
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1091 <SUZ>
A;Cross-references: GB:D78572; NID:g1545806; PIDN:BAA11416.1; PID:g1545807
C;Superfamily: leucine-rich alpha-2-glycoprotein repeat homology; proteoglycan amino-ter
F;36-61/Domain: proteoglycan amino-terminal homology <PAH>
F;71-94/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>
F;95-117/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>
F;118-141/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
F;142-165/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>
F;166-189/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>
F;191-213/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>
F;214-237/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>
F;238-261/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>
F;262-285/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>
F;286-309/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR10>
F;310-333/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR11>
F;334-357/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR12>
F;358-381/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR13>
F;385-408/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR14>
F;409-432/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR15>
F;440-485/Domain: proteoglycan carboxyl-terminal homology <PCH>


  Query Match             15.9%;  Score 402.5;  DB 2;  Length 1091;
  Best Local Similarity   24.1%;  Pred. No. 7.5e-20;
  Matches  134;  Conservative   77;  Mismatches  168;  Indels  177;  Gaps   14;

Qy      37 LNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIP 96
           |:|: |:|::|:  :|: |  ||:|:| ||:|  :  ||| ||: :: | |  | | :
Db     218 LDLNRNRIRLIEGLTFQGLDSLEVLRLQRNNISRLTDGAFWGLSKMHVLHLEYNSLVEVN 277

Qy      97 NGAFVYLSKLKELWLRNNPIESIPS-------------YAFNRIPSLRRLDLGELKRL-- 141
           :|:   |: | :| | || |  |                :|| :  |    | ||  |
Db     278 SGSLYGLTALHQLHLSNNSISRIQRDGWSFCQKLHELILSFNNLTRLDEESLAELSSLSI 337

Qy     142 --------SYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPG 193
                  |:|:||||:|| :|| |:                 ||  ::||        |
Db     338 LRLSHNAISHIAEGAFKGLKSLRVLD-----------------LDHNEISGTIEDT--SG 378

Qy     194 SFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIH 253
           :| || :| || :  ::|: : : ||  |:||  :||  | :  :  | |  : :|: ::
Db     379 AFTGLDNLSKLTLFGNKIKSVAKRAFSGLESLEHLNLGENAIRSVQFDAFAKMKNLKELY 438

Qy     254 LHHNPWNCNCDILWLSWWI--KDMAPSNTACCARCNTPPNLKGRYI-GELDQNY------ 304
           :    : |:| : ||  |:  : :    || ||    | :|||: |   |  ::
Db     439 ISSESFLCDCQLKWLPPWLMGRMLQAFVTATCAH---PESLKGQSIFSVLPDSFVCDDFP 495

Qy     305 -------------------FTCYA---------------------------------- 309
                             ||| |
Db     496 KPQIITQPETTMAVVGKDIRFTCSAASSSSSPMTFAWKKDNEVLANADMENFAHVRAQDG 555

Qy     310 ----------------------------------------------PVIVEPPADLNVTEG 324
                                                          |   : | |: :  |
Db     556 EVMEYTTILHLRHVTFGHEGRYQCIITNHFGSTYSHKARLTVNVLPSFTKIPHDIAIRTG 615

Qy     325 MAAELKCRASTSLT-SVSWITPNGTVMTHGAYKVRIAVLSDGTLNF-TNVTVQDTGMYTC 382
             | |:| |:      ::|    ||     | : |: |: |  : | |:| : | |:|:|
Db     616 TTARLECAATGHPNPQIAWQKDGGTDFP-AARERRMHVMPDDDVFFITDVKIDDMGVYSC 674

Qy     383 MVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNV--------GP 434
             || |: :|:||| |              |     | :|   |    |         |
Db     675 TAQNSAGSVSANATLTVLE-----------TPSLAVPLEDRVVTVGETVAFQCKATGSP 722

Qy     435 TPVVDWETTNVTTSLT 450
           || : |       |||
Db     723 TPRITWLKGGRPLSLT 738


RESULT   3
JC7763
neuronal leucine-rich repeat protein-3 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C;Accession: JC7763
R;Fukamachi, K.; Matsuoka, Y.; Kitanaka, C.; Kuchino, Y.; Tsuda, H.
Biochem. Biophys. Res. Commun. 287, 257-263, 2001
A;Title: Rat neuronal leucine-rich repeat protein-3: Cloning and regulation of the ge
A;Reference number: JC7763; PMID:11549284
A;Contents: Fibrosarcoma cells
A;Accession: JC7763
A;Molecule type: mRNA
A;Residues: 1-707 <FUK>
A;Cross-references: GB:AF291437
C;Comment: This protein, a new member of the neuronal leucine-rich repeat protein fam
in protein-protein interaction and functions as a cell adhesion molecule or soluble l
C;Genetics:
A;Gene: nlrr-3
C;Keywords: cell adhesion


  Query Match             15.7%;  Score 396;  DB 2;  Length 707;
  Best Local Similarity   25.9%;  Pred. No. 1.2e-19;
  Matches  127;  Conservative   79;  Mismatches  186;  Indels   98;  Gaps   13;

Qy       3 CPSVCSCS----------NQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIK--- 48
           || :|:|             : | | |    |   |  :  :|::| |  | | |:
Db      29 CPQLCTCEIRPWFTPRSIYMEASTVDCNDLGLLNFPARLPADTQILLLQTNNIARIEHST 88

Qy      49 ------------VNSFKHLRHLEILQLSR--------NHIRTIEIGAFNGLANLNTLELF 88
                        |:   : :: : ::|:        | :  :      ||:||  | :
Db      89 DFPPVNLTGLDLSQNNLSSVTNINVQKMSQLLSVYLEENKLTELPEKCLYGLSNLQELYVN 148

Qy      89 DNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGA 148
            | |: |  |||| |  |  | | :| :: | |  |  :|:|  | ||:   :  | :
Db     149 HNLLSAISPGAFVGLHNLLRLHLNSNRLQMINSKWFEALPNLEILMLGD-NPILRIKDMN 207

Qy     149 FEGLSNLRYLNLAMCNLREIP--------------------NLTPLIKLDE------LDL 182
           |: |  || | :|  || |:|                    |  | : | :      |||
Db     208 FQPLLKLRSLVIAGINLTEVPDDALVGLENLESISFYDNRLNKVPQVALQKAVNLKFLDL 267

Qy     183 SGNHLSAIRPGSFQGLMHLQKLWM--------IQS-----------------QIQVIER 216
           : | :: || | |  ::||::| :         | |                ::  |
Db     268 NKNPINRIRRGDFSNMLHLKELGINNMPELVSIDSLAVDNLPDLRKIEATNNPRLSYIHP 327

Qy     217 NAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSW---WIK 273
           |||  |  |  : |  | |: | |     | :|: | :| ||  |:| | |::     |:
Db     328 NAFFRLPKLESLMLNSNALSALYHGTIESLPNLKEISIHSNPIRCDCVIRWINMNKTNIR 387

Qy     274 DMAPSNTACCARCNTPPNLKGRYIGELD-QNYFTCYAPVIVEP--PADLNVTEGMAAELK 330
           | | :  |       ||  :|: : ::  ::     |:|     |: |:|       |
Db     388 FMEPDSLFCV----DPPEFQGQNVRQVHFRDMMEICLPLIAPESFPSILDVEADSYVSLH 443

Qy     331 CRASTS-LTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVG 389
           |||:      : ||||:|  :     : :  | |:|||:   :| :: |:|||: :| ||
Db     444 CRATAEPQPEIYWITPSGKRLLPNTLREKFYVHSEGTLDIRGITPKEGGLYTCIATNLVG 503

Qy     390 NTTASATLNV 399
               |  : |
Db     504 ADLKSIMIKV 513


RESULT   4
A53860
chondroadherin precursor - bovine
N;Alternate names: 38K leucine-rich protein
C;Species: Bos primigenius taurus (cattle)
C;Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 05-Nov-1999
C;Accession: A53860
```

```
R;Neame, P.J.; Sommarin, Y.; Boynton, R.E.; Heinegard, D.
J. Biol. Chem. 269, 21547-21554, 1994
A;Title: The structure of a 38-kDa leucine-rich protein (chondroadherin) isolated from b
A;Reference number: A53860; MUID:94342341
A;Accession: A53860
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-361 <NEA>
A;Cross-references: GB:U08018; NID:g470671; PIDN:AAA21330.1; PID:g470672
C;Superfamily: leucine-rich alpha-2-glycoprotein repeat homology; proteoglycan carboxyl-
C;Keywords: disulfide bond
F;300-346/Domain: proteoglycan carboxyl-terminal homology <PCH>


  Query Match             15.6%;  Score 394;  DB 2;  Length 361;
  Best Local Similarity   31.9%;  Pred. No. 6.9e-20;
  Matches  104;  Conservative   49;  Mismatches  139;  Indels   34;  Gaps    8;

Qy       3 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFK--------H 54
           ||  | | :    ||| :  |:::|  :|  |:|||| |   ::  |||:         |
Db      25 CPQNCHCHSDLQHVICDKVGLQKIPK-VSEKTKLLNLQRNNFPVLATNSFRAMPNLVSLH 83

Qy      55 LRHLEI----------------LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNG 98
           |:| :|                | || | || :  |||: |  |  | |  |::| :| |
Db      84 LQHCQIREVAAGAFRGLKQLIYLYLSHNDIRVLRAGAFDDLTELTYLYLDHNKVTELPRG 143

Qy      99 AFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYL 158
             |  | | || |  : | ||     || | | |   || :  || : : ||
Db     144 LLSPLVNLFILQLNNNKIRELRSGAFQGAKDLRWLYLSE-NSLSSLQPGALDDVENLAKF 202

Qy     159 NLAMCNLREIPN--LTPLIKLDELDLSGNHLSAIRPGSFQGL-MHLQKLWMIQSQIQVIE 215
           |    |   |:  |: |  ::|| || | | :|   :||    :|: ||:  : ::
Db     203 YLDRNQLSSYPSAALSKLRVVEELKLSHNPLKSIPDNAFQSFGRYLETLWLDNTNLEKFS 262

Qy     216 RNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDM 275
             ||  : :|  ::| :| |  || :   |   || : | :||| | | :  |  |::
Db     263 DGAFLGVTTLKHVHLENNRLHQLPSNF--PFDSLETLTLTNNPWKCTCQLRGLRRWLEAK 320

Qy     276 APSNTACCARCNTPPNLKGRYIGELD 301
             | ||   :|   :|::| : |
Db     321 TSRPDATCA---SPAKFRGQHIRDTD 343


RESULT   5
T42218
slit-1 protein homolog - rat
N;Alternate names: MEGF4 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C;Accession: T42218
R;Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A;Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A;Reference number: Z14126; MUID:98360089
A;Accession: T42218
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1531 <NAK>
A;Cross-references: EMBL:AB011530; NID:g3449289; PIDN:BAA32460.1; PID:g3449290
A;Experimental source: strain Sprague-Dawley; brain
C;Genetics:
A;Gene: MEGF4


  Query Match             15.4%;  Score 389.5;  DB 2;  Length 1531;
  Best Local Similarity   24.9%;  Pred. No. 9.3e-19;
  Matches  124;  Conservative   65;  Mismatches  146;  Indels  163;  Gaps   12;

Qy       3 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQ 62
           ||::|:|:   : | |    |: :|  |  ||  | |: | |  |  | |  |: | :||
Db      34 CPALCTCTG--TTVDCHGTGLQAIPKNIPRNTERLELNGNNITRIHKNDFAGLKQLRVLQ 91

Qy      63 LSRNHIRTIEIGAFN---------------------------------------------- 77
```

```
           | | |  :| |||:
Db      92 LMENQIGAVERGAFDDMKELERLRLNRNQLQVLPELLFQNNQALSRLDLSENSLQAVPRK 151

Qy      78 ---GLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLR--R 132
              |  :|  |:|  |::: |  |||  |  |: | | || | :||  :|| :| ||  |
Db     152 AFRGATDLKNLQLDKNQISCIEEGAFRALRGLEVLTLNNNNITTIPVSSFNHMPKLRTFR 211

Qy     133 LDLGEL---KRLSYISE--------GAF---EGLSNLRYLNL----------------- 160
           |    |      |:::|:         | |     | ::|| ||:
Db     212 LHSNHLFCDCHLAWLSQWLRQRPTIGLFTQCSGPASLRGLNVAEVQKSEFSCSGQGEAAQ 271

Qy     161 ------------AMCN--------------------------------LREIP--NLTP 173
                       |||:                                :: ||    :|
Db     272 VPACTLSSGSCPAMCSCSNGIVDCRGKGLTAIPANLPETMTEIRLELNGIKSIPPGAFSP 331

Qy     174 LIKLDELDLSGNHLSAIRPGSFQGLMH----------------------LQKLWMIQS 209
             ||  :||| | :: | | :||||                          || | :  :
Db     332 YRKLRRIDLSNNQIAEIAPDAFQGLRSLNSLVLYGNKITDLPRGVFGGLYTLQLLLLNAN 391

Qy     210 QIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLS 269
           :|  |  :|| :||:|  ::|  | :  |    || |  :: :||  ||: |:|:: ||:
Db     392 KINCIRPDAFQDLQNLSLLSLYDNKIQSLAKGTFTSLRAIQTLHLAQNPFICDCNLKWLA 451

Qy     270 WWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTC------YAPVIVEPPADLNVTE 323
             :::      |  ||| :|  |  : ||::    | |       : |   :   :   |
Db     452 DFLRTNPIETTG--ARCASPRRLANKRIGQIKSKKFRCSAKEQYFIPGTEDYHLNSECTS 509

Qy     324 GMAAELKCRASTSLTSVS 341
           :|   |||   |:   |
Db     510 DVACPHKCRCEASVVECS 527


RESULT   6
T13953
MEGF5 protein - rat
N;Alternate names: slit protein homolog
C;Species: Rattus norvegicus (Norway rat)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C;Accession: T13953
R;Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A;Title: Identification of high-molecular-weight proteins with multiple EGF-like moti
A;Reference number: Z14126; MUID:98360089
A;Accession: T13953
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1523 <NAK>
A;Cross-references: EMBL:AB011531; NID:g3449291; PIDN:BAA32461.1; PID:g3449292
C;Genetics:
A;Gene: MEGF5


  Query Match             14.5%;  Score 365.5;  DB 2;  Length 1523;
  Best Local Similarity   25.2%;  Pred. No. 4.2e-17;
  Matches  115;  Conservative   48;  Mismatches  137;  Indels  157;  Gaps    9;

Qy       3 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQ 62
           ||: |:||   : | |    || || ||  |   |:|  | |  |    |  |::| :|
Db      34 CPTKCTCS--AASVDCHGLGLRAVPRGIPRNAERLDLDRNNITRITKMDFTGLKNLRVLH 91

Qy      63 LSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSY 122
           |  | :  || |||  |  |  | |  |:|  :|   |    ||  | |  | |: ||
Db      92 LEDNQVSVIERGAFQDLKQLERLRLNKNKLQVLPELLFQSTPKLTRLDLSENQIQGIPRK 151

Qy     123 AFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNL---------------------- 160
           ||  :  :: | | :   :| | :|||  | :|  | |
Db     152 AFRGVTGVKNLQL-DNNHISCIEDGAFRALRDLEILTLNNNNISRILVTSFNHMPKIRTL 210

Qy     161 ------------------------------------------------------------ 160

Db     211 RLHSNHLYCDCHLAWLSDWLRQRRTIGQFTLCMAPVHLRGFSVADVQKKEYVCPGPHSEA 270

Qy     161 AMCN----------------------LREIP-NLTPLI---------------------- 175
```

```
Db    271 PACNANSLSCPSACSCSNNIVDCRGKGLTEIPANLPEGIVEIRLEQNSIKSIPAGAFIQY 330

Qy    176 -KLDELDLSGNHLSAIRPGSFQ-----------------------GLMHLQKLWMIQSQ 210
Db    331 KKLKRIDISKNQISDIAPDAFQGLKSLTSLVLYGNKITEIPKGLFDGLVSLQLLLLNANK 390

Qy    211 IQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSW 270
Db    391 INCLRVNTFQDLQNLNLLSLYDNKLQTISKGLFAPLQSIQTLHLAQNPFVCDCHLKWLAD 450

Qy    271 WIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTC 307
Db    451 YLQD-NPIETS-GARCSSPRRLANKRISQIKSKKFRC 485


RESULT   7
B36665
slit protein 2 precursor - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change 19-May-2000
C;Accession: B36665
R;Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.; Artavanis-Tsakonas, S.
Genes Dev. 4, 2169-2187, 1990
A;Title: slit: an extracellular protein necessary for development of midline glia and cc
A;Reference number: A36665; MUID:91099665
A;Accession: B36665
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1469 <ROT>
A;Cross-references: GB:X53959
C;Genetics:
A;Gene: FlyBase:sli
A;Cross-references: FlyBase:FBgn0003425
C;Superfamily: unassigned EGF-related proteins; EGF homology; leucine-rich alpha-2-glyco
F;66-91/Domain: proteoglycan amino-terminal homology <PAH1>
F;101-124/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>
F;125-148/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>
F;149-172/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
F;173-196/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>
F;197-220/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>
F;228-272/Domain: proteoglycan carboxyl-terminal homology <PCS1>
F;288-313/Domain: proteoglycan amino-terminal homology <PAH2>
F;323-346/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>
F;347-370/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>
F;371-394/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>
F;395-418/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>
F;419-442/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR10>
F;450-494/Domain: proteoglycan carboxyl-terminal homology <PCS2>
F;512-537/Domain: proteoglycan amino-terminal homology <PAH3>
F;547-571/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR11>
F;572-595/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR12>
F;596-619/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR13>
F;620-643/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR14>
F;651-695/Domain: proteoglycan carboxyl-terminal homology <PCS3>
F;708-733/Domain: proteoglycan amino-terminal homology <PAH4>
F;743-766/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR15>
F;767-790/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR16>
F;846-890/Domain: proteoglycan carboxyl-terminal homology <PCS4>
F;1028-1061/Domain: EGF homology <EGF>
F;1068-1099/Domain: EGF homology <EGF2>
F;1115-1148/Domain: EGF homology <EGF1>


  Query Match             13.9%;  Score 350;  DB 2;  Length 1469;
  Best Local Similarity   23.9%;  Pred. No. 4.8e-16;
  Matches  109;  Conservative   65;  Mismatches  152;  Indels  130;  Gaps    9;

Qy      2 TCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEIL 61
Db    294 SCPHPCRCADGI--VDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKSFSSFRRLRRI 351

Qy     62 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 121
```

```
Db    352 DLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLLNANEISCIRK 411

Qy    122 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM------CNLR--------- 166
Db    412 DAFRDLHSLSLLSLYD-NNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRWLADYLHKN 470

Qy    167 ------------------------------------------------------------ 166
Db    471 PIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKLSGECRMDSDCPAMCHCEGTTV 530

Qy    167 --------EIPNLTPL--------------IKLD----------ELDLSGNHLSAIRPGS 194
Db    531 DCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKRNQLTGIEPNA 590

Qy    195 FQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 254
Db    591 FEGASHIQELQLGENKIKEISNKMFLGLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNL 650

Qy    255 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVE 314
Db    651 ASNPFNCNCHLAWFAECVRKKSLNGGA--ARCGAPSKVRDVQIKDLPHSEFKCSS----- 703

Qy    315 PPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVM 350
Db    704 -----ENSEGCLGDGYCPPSCTCT--------GTVV 726


RESULT   8
A36665
slit protein 1 precursor - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 30-Apr-1991 #sequence_revision 30-Apr-1991 #text_change 17-Nov-2000
C;Accession: A36665; A31640; S13523
R;Rothberg, J.M.; Jacobs, J.R.; Goodman, C.S.; Artavanis-Tsakonas, S.
Genes Dev. 4, 2169-2187, 1990
A;Title: slit: an extracellular protein necessary for development of midline glia and
A;Reference number: A36665; MUID:91099665
A;Accession: A36665
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1480 <ROT>
A;Cross-references: GB:X53959; NID:g8614; PIDN:CAA37910.1; PID:g8615
R;Rothberg, J.M.; Hartley, D.A.; Walther, Z.; Artavanis-Tsakonas, S.
Cell 55, 1047-1059, 1988
A;Title: slit: An EGF-homologous locus of D. melanogaster involved in the development
A;Reference number: A31640; MUID:89077533
A;Accession: A31640
A;Molecule type: DNA
A;Residues: 881-1182,'G',1185-1404,'GT',1463-1464,'YHA' <RO2>
A;Cross-references: GB:M23543; NID:g340939; PID:g514357
C;Genetics:
A;Gene: FlyBase:sli
A;Cross-references: FlyBase:FBgn0003425
A;Introns: 1351/3
C;Superfamily: unassigned EGF-related proteins; EGF homology; leucine-rich alpha-2-gl
C;Keywords: alternative splicing; growth factor
F;66-91/Domain: proteoglycan amino-terminal homology <PAH1>
F;101-124/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>
F;125-148/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>
F;149-172/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
F;173-196/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>
F;197-220/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>
F;228-272/Domain: proteoglycan carboxyl-terminal homology <PCS1>
F;288-313/Domain: proteoglycan amino-terminal homology <PAH2>
F;323-346/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>
F;347-370/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>
F;371-394/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>
F;395-418/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>
F;419-442/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR10>
F;450-494/Domain: proteoglycan carboxyl-terminal homology <PCS2>
F;512-537/Domain: proteoglycan amino-terminal homology <PAH3>
F;547-571/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR11>
```

```
F;572-595/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR12>
F;596-619/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR13>
F;620-643/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR14>
F;651-695/Domain: proteoglycan carboxyl-terminal homology <PCS3>
F;708-733/Domain: proteoglycan amino-terminal homology <PAH4>
F;743-766/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR15>
F;767-790/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR16>
F;791-814/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR17>
F;815-838/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR18>
F;846-890/Domain: proteoglycan carboxyl-terminal homology <PCS4>
F;1028-1061/Domain: EGF homology <EGF>
F;1068-1099/Domain: EGF homology <EGF2>
F;1115-1148/Domain: EGF homology <EGF1>


  Query Match             13.9%;  Score 350;  DB 2;  Length 1480;
  Best Local Similarity   23.9%;  Pred. No. 4.8e-16;
  Matches  109;  Conservative   65;  Mismatches  152;  Indels  130;  Gaps    9;

Qy       2 TCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEIL 61
           :||  | |::    | |  |:|  ||  :  :|  : | :| |  :   ||   | |  :
Db     294 SCPHPCRCADGI--VDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKSFSSFRRLRRI 351

Qy      62 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 121
             || |:|  |   | :||  | || |: |::  :|:| |  |  |: | |  | |  |
Db     352 DLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLLNANEISCIRK 411

Qy     122 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM------CNLR--------- 166
             ||  : ||  | | :   :  :: | |: : ::: ::||       ||||
Db     412 DAFRDLHSLSLLSLYD-NNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRWLADYLHKN 470

Qy     167 ------------------------------------------------------------ 166

Db     471 PIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKLSGECRMDSDCPAMCHCEGTTV 530

Qy     167 --------EIPNLTPL--------------IKLD----------ELDLSGNHLSAIRPGS 194
                   |||   ||              |  |          :|:|  | |: | | :
Db     531 DCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKRNQLTGIEPNA 590

Qy     195 FQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 254
           |:|  |:|:|  : :::|: |    |  |  |  :|| | :: :    |  |: |  ::|
Db     591 FEGASHIQELQLGENKIKEISNKMFLGLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNL 650

Qy     255 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVE 314
             ||:|||| : | :  ::  : :  |  |||  |  ::   | :|  : | | :
Db     651 ASNPFNCNCHLAWFAECVRKKSLNGGA--ARCGAPSKVRDVQIKDLPHSEFKCSS----- 703

Qy     315 PPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVM 350
                 :||   :  |  | : |        |||:
Db     704 -----ENSEGCLGDGYCPPSCTCT--------GTVV 726


RESULT   9
T28714
hypothetical protein T21D12.9a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T28714
R;Woessner, J.
submitted to the EMBL Data Library, August 1997
A;Description: The sequence of C. elegans cosmid T21D12.
A;Reference number: Z20514
A;Accession: T28714
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-789 <WOE>
A;Cross-references: EMBL:AF016687; PIDN:AAC48096.1; GSPDB:GN00022; CESP:T21D12.9a
A;Experimental source: strain Bristol N2; clone T21D12
C;Genetics:
A;Gene: CESP:T21D12.9a
A;Map position: 4
A;Introns: 38/2; 84/2; 132/2; 204/2; 275/2; 351/2; 519/3; 615/1; 714/1; 758/2


  Query Match             13.7%;  Score 345;  DB 2;  Length 789;
  Best Local Similarity   26.3%;  Pred. No. 4.7e-16;
  Matches  109;  Conservative   70;  Mismatches  177;  Indels   58;  Gaps   12;

Qy      31 STNTRL-LNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFD 89
           | || : | |  | |  :   ||  || || | |:|| || :   ||| | :|  : |
Db     101 SFNTLVTLKLARNHITTLNQFSFSRLRKLESLDLTRNMIREVRFLAFNQLPSLQNVSLAR 160

Qy      90 NRLTTIPNGAFVYLSKLKELWLRNNPIESIP------------SYAFNRI--------- 127
           | :  : :| |     || | |  | ::::                ::|:|
Db     161 NDVYRLDDGMFYACEGLKHLNLSTNRVQAVTEGWMFGLTSLEVLDLSYNQIQSFHISSWS 220

Qy     128 --PSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPN--LTPLIKLDELDLS 183
             | |: | |    |:  :  |:|  |  |  | |:  ::  :    |  :  | :|||
Db     221 HTPKLKWLSL-HSNRIQSLPSGSFRVLRQLEELILSANSIDSLHKFALVGMSSLHKLDLS 279

Qy     184 GNHLS-AIRPGSF---QGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLP 239
           | |:  :  |:      :  |: |    :|::|| : ||:   :| |::|  | :  :
Db     280 SNTLAVCVEDGAVLYNTSMPFLRSLRFTNNQLRVIPKRAFERFPALEELDLTDNPIATIH 339

Qy     240 HDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGE 299
           : | ||  |:|: :: :   |:| | ||: ||  :    ::  |:|: || |   |:
Db     340 PEAFEPL-ELKRLVMNSSSILCDCQISWLASWIYRLKLDKSSIIAKCSYPPPLADLYVVA 398

Qy     300 LDQNYFTCY----APVIVEPPADLNVTEGMAAELKCRA-STSLTSVSW------------ 342
           :|    ||:       ||  | :::   |  |   |      |  |: |
Db     399 IDTANLTCHNDSPRAKIVRQPVEVSTLIGEKARFTCNVYGASPLSIEWRVMENGQPRVLV 458

Qy     343 ------ITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGN 390
                 :: | | : :| :  |   |:   |   || : |   | |:  |  |:
Db     459 QDSATFLSINRTAVVNGTFDER--ELAAAELLLDNVAMTDNSEYQCVARNRFGS 510


RESULT  10
T28715
hypothetical protein T21D12.9b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T28715
R;Woessner, J.
submitted to the EMBL Data Library, August 1997
A;Description: The sequence of C. elegans cosmid T21D12.
A;Reference number: Z20514
A;Accession: T28715
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1355 <WOE>
A;Cross-references: EMBL:AF016687; PIDN:AAC48095.1; GSPDB:GN00022; CESP:T21D12.9b
A;Experimental source: strain Bristol N2; clone T21D12
C;Genetics:
A;Gene: CESP:T21D12.9b
A;Map position: 4
A;Introns: 38/2; 84/2; 132/2; 204/2; 275/2; 351/2; 519/3; 615/1; 714/1; 758/2; 786/3;


  Query Match             13.7%;  Score 345;  DB 2;  Length 1355;
  Best Local Similarity   26.3%;  Pred. No. 9.5e-16;
  Matches  109;  Conservative   70;  Mismatches  177;  Indels   58;  Gaps   12;

Qy      31 STNTRL-LNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFD 89
           | || : | |  | |  :   ||  || || | |:|| || :   ||| | :|  : |
Db     101 SFNTLVTLKLARNHITTLNQFSFSRLRKLESLDLTRNMIREVRFLAFNQLPSLQNVSLAR 160

Qy      90 NRLTTIPNGAFVYLSKLKELWLRNNPIESIP------------SYAFNRI--------- 127
           | :  : :| |     || | |  | ::::                ::|:|
Db     161 NDVYRLDDGMFYACEGLKHLNLSTNRVQAVTEGWMFGLTSLEVLDLSYNQIQSFHISSWS 220

Qy     128 --PSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPN--LTPLIKLDELDLS 183
             | |: | |    |:  :  |:|  |  |  | |:  ::  :    |  :  | :|||
Db     221 HTPKLKWLSL-HSNRIQSLPSGSFRVLRQLEELILSANSIDSLHKFALVGMSSLHKLDLS 279

Qy     184 GNHLS-AIRPGSF---QGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLP 239
```

```
Db    280 SNTLAVCVEDGAVLYNTSMPFLRSLRFTNNQLRVIPKRAFERFPALEELDLTDNPIATIH 339

Qy    240 HDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGE 299

Db    340 PEAFEPL-ELKRLVMNSSSILCDCQISWLASWIYRLKLDKSSIIAKCSYPPPLADLYVVA 398

Qy    300 LDQNYFTCY----APVIVEPPADLNVTEGMAAELKCRA-STSLTSVSW---------- 342

Db    399 IDTANLTCHNDSPRAKIVRQPVEVSTLIGEKARFTCNVYGASPLSIEWRVMENGQPRVLV 458

Qy    343 ------ITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGN 390

Db    459 QDSATFLSINRTAVVNGTFDER--ELAAAELLLDNVAMTDNSEYQCVARNRFGS 510


RESULT 11
T42626
secreted leucine-rich repeat-containing protein SLIT2 - mouse (fragment)
N;Alternate names: neurogenic extracellular slit protein
C;Species: Mus musculus (house mouse)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C;Accession: T42626
R;Holmes, G.P.; Negus, K.; Burridge, L.; Raman, S.; Algar, E.; Yamada, T.; Little, M.H.
Mech. Dev. 79, 57-72, 1998
A;Title: Distinct but overlapping expression patterns of two vertebrate slit homologs im
A;Reference number: Z22177; MUID:99279238
A;Accession: T42626
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1025 <HOL>
A;Cross-references: EMBL:AF074960; NID:g4151258; PID:g4151259; PIDN:AAD04345.1
C;Genetics:
A;Gene: Slit2

  Query Match             13.6%;  Score 344.5;  DB 2;  Length 1025;
  Best Local Similarity   24.3%;  Pred. No. 7.1e-16;
  Matches  102;  Conservative   62;  Mismatches  149;  Indels  107;  Gaps    8;

Qy      3 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNS-FKHLRHLEIL 61

Db      2 CPEKCRCEG--TTVDCSNQRLNKIPDHIPQYTAELRLNNNEFTVLEATGIFKKLPQLRXI 59

Qy     62 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 121

Db     60 NFSNNKITDIEEGAFEGASGVNEILLTSNRLENVQHKMFKGLESLKTLMLRSNRISCVGN 119

Qy    122 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM----CN------------ 164

Db    120 DSFIGLGSVRLLSLYD-NQITTVAPGAFDXLHSLSTLNLLANPFNCNCHLAWLGEWLRRK 178

Qy    165 -------------LREIP-----------------NLTPLIK---------------- 176

Db    179 RIVTGNPRCQKPYFLKEIPIQDVAIQDFTCDDGNDDNSCSPLSRCPSECTCLDTXVRCSN 238

Qy    177 ----------------------------------LDELDLSGNHLSAIRPGSFQGLMH 200

Db    239 KGLKVLPKGIPKDVTELYLDGNQFTLVPKELSNYKHLTLIDLSNNRISTLSNQXFSNMTQ 298

Qy    201 LQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWN 260

Db    299 LLTLILSYNRLRCIPPRTFDGLKSLRLLSLHGNDISVVPEGAFNDLSALSHLAIGANPLY 358

Qy    261 CNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLN 320

Db    359 CDCNMQWLSDWVK--SEYKEPGIARCAGPGEMADKLLLTTPSKKFTCQGPMDITIQAKCN 416


RESULT 12
A60164
platelet membrane glycoprotein V precursor - human
C;Species: Homo sapiens (man)
C;Date: 12-Jan-1993 #sequence_revision 24-Feb-1994 #text_change 05-Nov-1999
C;Accession: A48030; A60164; A35483; B35483; C35483; A60432; A47507; S34329
R;Lanza, F.; Morales, M.; de La Salle, C.; Cazenave, J.P.; Clemetson, K.J.; Shimomura
J. Biol. Chem. 268, 20801-20807, 1993
A;Title: Cloning and characterization of the gene encoding the human platelet glycopr
A;Reference number: A48030; MUID:94012616
A;Accession: A48030
A;Molecule type: DNA
A;Residues: 1-560 <LA2>
A;Cross-references: EMBL:Z23091; NID:g312501; PIDN:CAA80637.1; PID:g312502
R;Shimomura, T.; Fujimura, K.; Maehama, S.; Takemoto, M.; Oda, K.; Fujimoto, T.; Oyam
Blood 75, 2349-2356, 1990
A;Title: Rapid purification and characterization of human platelet glycoprotein V: th
A;Reference number: A60164; MUID:90275263
A;Accession: A60164
A;Molecule type: protein
A;Residues: 365-384,'X',386-390,'X',392-395,'X',397;188-208,'I',210;27-50,'X',52-53;1
,'XX',108,'T';61-72,'TK',75-77;'V',56-57;'G',479-487,'X',489-498,'X',500,'X',502-503,
R;Roth, G.J.; Church, T.A.; McMullen, B.A.; Williams, S.A.
Biochem. Biophys. Res. Commun. 170, 153-161, 1990
A;Title: Human platelet glycoprotein V: a surface leucine-rich glycoprotein related t
A;Reference number: A35483; MUID:90321220
A;Accession: A35483
A;Molecule type: protein
A;Residues: 145-166,'I',168-169,'X',171-172 <ROT>
A;Note: this proteolytic fragment was designated peptide M392
A;Accession: B35483
A;Molecule type: protein
A;Residues: 121-129,'W',131-135;466-468,'X',470 <RO2>
A;Note: this material was designated peptide M393 but may contain two peptides
A;Accession: C35483
A;Molecule type: protein
A;Residues: 252-266,'H',268-272,'X',274-279,'I',281-284,'I',286 <RO3>
A;Note: this proteolytic fragment was designated peptide M401
R;Zafar, R.S.; Walz, D.A.
Thromb. Res. 53, 31-44, 1989
A;Title: Platelet membrane glycoprotein V: characterization of the thrombin-sensitive
A;Reference number: A60432; MUID:89162331
A;Accession: A60432
A;Molecule type: protein
A;Residues: 477-478,'FX',481-485,'E',487,'V',489-492,'NQ',495,'E',497-498 <ZAF>
R;Hickey, M.J.; Hagen, F.S.; Yagi, M.; Roth, G.J.
Proc. Natl. Acad. Sci. U.S.A. 90, 8327-8331, 1993
A;Title: Human platelet glycoprotein V: characterization of the polypeptide and the r
A;Reference number: A47507; MUID:93391348
A;Accession: A47507
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-560 <RES>
A;Cross-references: GB:L11238; NID:g388759; PIDN:AAA03069.1; PID:g388760
C;Comment: This platelet membrane protein is a substrate for thrombin.
C;Comment: The amino end of the intact protein is blocked.
C;Comment: This protein is absent in Bernard-Soulier syndrome.
C;Genetics:
A;Gene: GDB:GP5
A;Cross-references: GDB:230236; OMIM:173511
A;Map position: 5pter-5qter
C;Superfamily: leucine-rich alpha-2-glycoprotein repeat homology
C;Keywords: blocked amino end; glycoprotein; platelet; tandem repeat; transmembrane p

  Query Match             13.1%;  Score 329.5;  DB 2;  Length 560;
  Best Local Similarity   29.7%;  Pred. No. 3.5e-15;
  Matches   83;  Conservative   43;  Mismatches  104;  Indels   49;  Gaps    3;

Qy     22 NLREVPDG-----------------ISTNTRLLN---------LHENQIQIIKVNSFKH 54

Db    181 NLTHLPKGLLGAQAKLERLLLHSNRLVSLDSGLLNSLGALTELQFHRNHIRSIAPGAFDR 240

Qy     55 LRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNN 114

Db    241 LPNLSSLTLSRNHLAFLPSALFLHSHNLTLLTLFENPLAELPGVLFGEMGGLQELWLNRT 300
```

```
Qy    115 PIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPL 174
Db    301 QLRTLPAAAFRNLSRLRYLGVTLSPRLSALPQGAFQGLGELQV---------------- 343

Qy    175 IKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNN 234
Db    344 -----LALHSNGLTALPDGLLRGLGKLRQVSLRRNRLRALPRALFRNLSSLESVQLDHNQ 398

Qy    235 LTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIK 273
Db    399 LETLPGDVFGALPRLTEVLLGHNSWRCDCGLGPFLGWLR 437


RESULT  13
JC5239
insulin-like growth factor acid-labile chain - baboon
C;Species: Papio sp. (baboon)
C;Date: 17-Apr-1997 #sequence_revision 09-May-1997 #text_change 09-May-1997
C;Accession: JC5239
R;Delhanty, P.; Baxter, R.C.
Biochem. Biophys. Res. Commun. 227, 897-902, 1996
A;Title: The cloning and expression of the baboon acid-labile subunit of the insulin-lik
A;Reference number: JC5239; MUID:97040714
A;Contents: liver
A;Accession: JC5239
A;Molecule type: mRNA
A;Residues: 1-605 <DEL>
C;Comment: This factor is structurally related to proinsulin and have insuline-like meta


  Query Match             13.1%;  Score 329.5;  DB 2;  Length 605;
  Best Local Similarity   21.0%;  Pred. No. 3.9e-15;
  Matches  119;  Conservative   61;  Mismatches  118;  Indels  269;  Gaps   12;

Qy      3 CPSVCSCS-----NQFSKVICVRKNLREVPDGISTNTRL---------------------- 36
Db     41 CPATCACSYDDEVNELS-VFCSSRNLTRLPDGIPGGTQALWLDSNNLSSIPPAAFRNLSS 99

Qy     37 --------------------------LNLHENQIQIIKVNSFKHLRHLEILQLSRNHIR 69
Db    100 LAFLNLQGGQLGSLEPQALLGLENLCHLHLERNQLRSLAVGTFAYTPALALLGLSNNRLS 159

Qy     70 TIEIGAFNGLANLNTLELFDNRLTTIPNGAF-----------------VY--------LSK 105
Db    160 RLEDGLFEGLGNLWDLNLGWNSLAVLPDAAFRGLGGLRELVLAGNRLAYLQPALFSGLAE 219

Qy    106 LKELWLRNNPIESIPSYAFNRIPSLRRL------------------------------- 133
Db    220 LRELDLSRNALRAIKANVFAQLPRLQKLYLDRNLIAAVAPGAFLGLKALRWLDLSHNRVA 279

Qy    134 ------------------------------------------------------DLGE 137
Db    280 GLLEDTFPGLLGLRVLRLSHNAIASLRPRTFEDLHFLEELQLGHNRIRQLAERSFEGLGQ 339

Qy    138 LK-------RLSYISEGAFEGLSNLRYLNLAMCNLREIPN--LTPLIKLDELDLSGNHLS 188
Db    340 LEVLTLDHNQLQEVKVGAFLGLTNVAVMNLSGNCLRNLPEQVFRGLGKLHSLHLEGSCLG 399

Qy    189 AIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPL-- 246
Db    400 RIRPHTFAGLSGLRRLFLKDNGLVGIEEQSLWGLAELLELDLTSNQLTHLPHQLFQGLGK 459

Qy    247 -------------------------------HH--------------------------- 248
Db    460 LEYLLLSHNRLAELPADALGPLQRAFWLDVSHNRLEALPGSLLASLGRLRYLNLRNNSLR 519

Qy    249 --------LERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCAR-------------- 285
Db    520 TFTPQPPGLERLWLEGNPWDCSCPLKAL----RDFALQNPSAVPRFVQAICEGDDCQPPV 575

Qy    286 -------CNTPPNLKGRYIGELDQNYF 305
Db    576 YTYNNITCASPPEVAGLDLRDLGEAHF 602


RESULT  14
A41915
insulin-like growth factor-binding complex acid-labile chain precursor - human
N;Alternate names: Acid-Labile Subunit (ALS)
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Nov-1999
C;Accession: A41915
R;Leong, S.R.; Baxter, R.C.; Camerato, T.; Dai, J.; Wood, W.I.
Mol. Endocrinol. 6, 870-876, 1992
A;Title: Structure and functional expression of the acid-labile subunit of the insuli
A;Reference number: A41915; MUID:92357025
A;Accession: A41915
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-605 <LEO>
A;Cross-references: GB:M86826; NID:g184807; PIDN:AAA36047.1; PID:g184808
A;Experimental source: liver
A;Note: sequence extracted from NCBI backbone (NCBIP:110171)
C;Superfamily: leucine-rich alpha-2-glycoprotein repeat homology
F;75-98/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>
F;99-122/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>
F;123-146/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
F;147-170/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>
F;171-194/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>
F;195-218/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>
F;219-242/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>
F;243-266/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>
F;267-290/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>
F;291-314/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR10>
F;315-338/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR11>
F;339-362/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR12>
F;363-386/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR13>
F;387-410/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR14>
F;411-434/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR15>
F;435-458/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR16>
F;459-482/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR17>
F;483-506/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR18>
F;507-529/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR19>


  Query Match             12.5%;  Score 314.5;  DB 2;  Length 605;
  Best Local Similarity   31.1%;  Pred. No. 4.3e-14;
  Matches   89;  Conservative   47;  Mismatches  117;  Indels   33;  Gaps    5;

Qy      3 CPSVCSCS-----NQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRH 57
Db     41 CPAACVCSYDDDADELS-VFCSSRNLTRLPDGVPGGTQALWLDGNNLSSVPPAAFQNLSS 99

Qy     58 LEI-----------------------LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLT 93
Db    100 LGFLNLQGGQLGSLEPQALLGLENLCHLHLERNQLRSLALGTFAHTPALASLGLSNNRLS 159

Qy     94 TIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLS 153
Db    160 RLEDGLFEGLGSLWDLNLGWNSLAVLPDAAFRGLGSLRELVLAG-NRLAYLQPALFSGLA 218

Qy    154 NLRYLNLAMCNLREIPN--LTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQI 211
Db    219 ELRELDLSRNALRAIKANVFVQLPRLQKLYLDRNLIAAVAPGAFLGLKALRWLDLSHNRV 278

Qy    212 QVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHN 257
Db    279 AGLLEDTFPGLLGLRVLRLSHNAIASLRPRTFKDLHFLEELQLGHN 324


RESULT  15
S46224
peroxidasin - fruit fly (Drosophila sp.)
C;Species: Drosophila sp.
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Oct-2000
C;Accession: S46224
```

```
R;Nelson, R.E.; Fessler, L.I.; Takagi, Y.; Blumberg, B.; Keene, D.R.; Olson, P.F.; Parke
EMBO J. 13, 3438-3447, 1994
A;Title: Peroxidasin: a novel enzyme-matrix protein of Drosophila development.
A;Reference number: S46224; MUID:94341255
A;Accession: S46224
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1535 <NEL>
A;Cross-references: GB:U11052; NID:g531384; PIDN:AAA61568.1; PID:g531385
C;Superfamily: peroxidasin; myeloperoxidase homology; proteoglycan amino-terminal homolo
F;19-44/Domain: proteoglycan amino-terminal homology <PAH4>
F;661-1350/Domain: myeloperoxidase homology <MPX>


  Query Match            12.4%;  Score 313;  DB 2;  Length 1535;
  Best Local Similarity  24.7%;  Pred. No. 1.8e-13;
  Matches   97;  Conservative   52;  Mismatches  136;  Indels  108;  Gaps   12;

Qy       3 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQ 62
           ||: |:|  :   | |:|  |  ||                          |  :  : |
Db      26 CPAGCTCLER--TVRCIRAKLSAVP-----------------------KLPQDTQTLD 58

Qy      63 LSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSY 122
           |  |||  :   ||:||| | || | || | |  : :||   |: |: ::| || :  :|:
Db      59 LRFNHIEELPANAFSGLAQLTTLFLNDNELAYLQDGALNGLTALRFVYLNNNRLSRLPAT 118

Qy     123 AFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDL 182
            | |:| |               || |  | |
Db     119 IFQRMPRL--------------EGIF--LEN--------------------------- 133

Qy     183 SGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDL 242
                             :| : : :       ||||  |  : : :| || || |
Db     134 ------------------NDIWQLPAGL-------FDNLPRLNRLIMYNNKLTQLPVDG 167

Qy     243 FTPLHHLERIHLHHN--PWNCNCDI--LWLSWWIKDMAPSNTACCARCNTPPNLKGRYIG 298
           |  |::|:|: |  |    :|| :  ||  | : |:     :    |  |  |: :
Db     168 FNRLNNLKRLRLDGNAIDIDCNCGVYSLWRRWHL-DVQRQLVSISLTCAAPQMLQNQGFS 226

Qy     299 ELDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSL--TSVSWITPNGTVMTHGAYK 356
            | :::| |  |  :  | |  |  |  ||  |   | |    ::|:  : |       :
Db     227 SLGEHHFKCAKPQFLVAPQDAQVAAGEQVELSCEV-TGLHRPQITWM--HNTQELGLEEQ 283

Qy     357 VRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVG 389
            :  :|  |:|   :    | |:| |:  | :|
Db     284 TQAEILPSGSLLHRSADTSDMGIYQCIARNEMG 316


Search completed: August 26, 2002, 15:47:05
Job time: 147 sec
```

OM protein - protein search, using sw model

Run on: August 26, 2002, 15:46:43 ; Search time 13.44 Seconds
(without alignments)
1377.079 Million cell updates/sec

Title: US-09-905-056-292_COPY_45_522
Perfect score: 2524
Sequence: 1 QTCPSVCSCSNQFSKVICVR.....KTFTIPVTDINSGIPGIDEV 478

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	430.5	17.1	713	1	GAC1_HUMAN	O75325 homo sapien
2	394	15.6	361	1	CHAD_BOVIN	Q27972 bos taurus
3	367	14.5	567	1	GPV_MOUSE	O08742 mus musculu
4	356	14.1	567	1	GPV_RAT	O08770 rattus norv
5	350	13.9	1480	1	SLIT_DROME	P24014 drosophila
6	329.5	13.1	560	1	GPV_HUMAN	P40197 homo sapien
7	329.5	13.1	605	1	ALS_PAPHA	O02833 papio hamad
8	314.5	12.5	605	1	ALS_HUMAN	P35858 homo sapien
9	309	12.2	603	1	ALS_RAT	P35859 rattus norv
10	302.5	12.0	603	1	ALS_MOUSE	P70389 mus musculu
11	302	12.0	359	1	PGS2_HUMAN	P07585 homo sapien
12	295	11.7	360	1	PGS2_PIG	Q9xsd9 sus scrofa
13	294	11.6	360	1	PGS2_HORSE	O46542 equus cabal
14	293	11.6	360	1	PGS2_BOVIN	P21793 bos taurus
15	291	11.5	357	1	PGS2_CHICK	P28675 gallus gall
16	290	11.5	360	1	PGS2_CANFA	Q29393 canis famil
17	289	11.5	360	1	PGS2_RABIT	Q28888 oryctolagus
18	286.5	11.4	536	1	CBP8_HUMAN	P22792 homo sapien
19	286	11.3	354	1	PGS2_MOUSE	P28654 mus musculu
20	283	11.2	354	1	PGS2_RAT	Q01129 rattus norv
21	282	11.2	626	1	GPBA_HUMAN	P07359 homo sapien
22	276	10.9	353	1	KERA_COTJA	Q9de66 coturnix co
23	274	10.9	353	1	KERA_CHICK	O42235 gallus gall
24	273.5	10.8	331	1	PLIB_AGKBL	O93233 agkistrodon
25	271.5	10.8	368	1	PGS1_HUMAN	P21810 homo sapien
26	271.5	10.8	369	1	PGS1_CANFA	O02678 canis famil
27	271.5	10.8	372	1	PGS1_HORSE	O46403 equus cabal
28	270.5	10.7	369	1	PGS1_MOUSE	P28653 mus musculu
29	270.5	10.7	369	1	PGS1_RAT	P47853 rattus norv
30	269.5	10.7	369	1	PGS1_SHEEP	O46390 ovis aries
31	268.5	10.6	369	1	PGS1_BOVIN	P21809 bos taurus
32	268	10.6	423	1	OMD_RAT	Q9z1s7 rattus norv
33	259.5	10.3	966	1	Y918_HUMAN	O94991 homo sapien
34	254.5	10.1	1049	1	TLR7_HUMAN	Q9nyk1 homo sapien
35	254	10.1	905	1	TLR3_MOUSE	Q99mb1 mus musculu
36	253	10.0	977	1	Y848_HUMAN	O94933 homo sapien
37	250.5	9.9	382	1	PRLP_HUMAN	P51888 homo sapien
38	249	9.9	904	1	TLR3_HUMAN	O15455 homo sapien
39	248	9.8	352	1	KERA_BOVIN	O62702 bos taurus
40	247	9.8	423	1	OMD_MOUSE	O35103 mus musculu
41	246.5	9.8	1315	1	CHAO_DROME	P12024 drosophila
42	246	9.7	352	1	KERA_HUMAN	O60938 homo sapien
43	245.5	9.7	440	1	OMGP_HUMAN	P23515 homo sapien
44	245.5	9.7	782	1	CHAO_TRICA	P82963 tribolium c
45	245	9.7	440	1	OMGP_MOUSE	Q63912 mus musculu

ALIGNMENTS

RESULT 1
GAC1_HUMAN
ID GAC1_HUMAN STANDARD; PRT; 713 AA.
AC O75325;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Glioma amplified on chromosome 1 protein precursor.
GN GAC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Glial tumor;
RX MEDLINE=98324709; PubMed=9662332;
RA Malfoy B., Almeida A., Zhu X.X., Vogt N., Tyagi R., Muleris M.,
RA Dutrillaux A.-M., Dutrillaux B., Ross D., Hanash S.;
RT "GAC1, a new member of the leucine-rich repeat superfamily on
RT chromosome band 1q32.1, is amplified and overexpressed in malignant
RT gliomas.";
RL Oncogene 16:2997-3002(1998).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -!- TISSUE SPECIFICITY: OVERAMPLIFIED IN MALIGNANT GLIOMAS.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 11 LEUCINE-RICH REPEATS (LRR).
CC ----------

CC ----------
DR EMBL; AF030435; AAC39792.1; -.
DR MIM; 605492; -.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR003591; LRR_typ.
DR Pfam; PF00047; ig; 1.
DR Pfam; PF00560; LRR; 10.
DR Pfam; PF01463; LRRCT; 1.
DR Pfam; PF01462; LRRNT; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00408; IGc2; 1.
DR SMART; SM00370; LRR; 6.
DR SMART; SM00082; LRRCT; 1.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_TYP; 2.

```
KW   Immunoglobulin domain; Transmembrane; Glycoprotein; Repeat;
KW   Leucine-rich repeat; Signal.
FT   SIGNAL       1    18       POTENTIAL.
FT   CHAIN       19   713       GLIOMA AMPLIFIED ON CHROMOSOME 1 PROTEIN.
FT   DOMAIN      19   630       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   631   651       POTENTIAL.
FT   DOMAIN     652   713       CYTOPLASMIC (POTENTIAL).
FT   REPEAT      92   115       LRR 1.
FT   REPEAT     116   139       LRR 2.
FT   REPEAT     140   163       LRR 3.
FT   REPEAT     165   187       LRR 4.
FT   REPEAT     188   211       LRR 5.
FT   REPEAT     213   235       LRR 6.
FT   REPEAT     236   259       LRR 7.
FT   REPEAT     261   283       LRR 8.
FT   REPEAT     309   333       LRR 9.
FT   REPEAT     334   357       LRR 10.
FT   REPEAT     359   385       LRR 11.
FT   DOMAIN     438   504       IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID   445   497       BY SIMILARITY.
FT   CARBOHYD    94    94       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   381   381       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   555   555       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   583   583       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   713 AA;  78798 MW;  EC8BC0FD240C9396 CRC64;

  Query Match             17.1%;  Score 430.5;  DB 1;  Length 713;
  Best Local Similarity   25.1%;  Pred. No. 2.9e-23;
  Matches  140;  Conservative   83;  Mismatches  211;  Indels  123;  Gaps   12;

Qy      3 CPSVCSCS-----------NQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNS 51
          ||  |:|             : : | |    |  ||  :   |: | |  | |  :  :
Db     29 CPPQCACQIRPWYTPRSSYREATTVDCNDLFLTAVPPALPAGTQTLLLQSNSIVRVDQSE 88

Qy     52 FKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWL 111
            :| :|  | ||:|         |: |  | :| | :|:|| : : :|  |: |:||:|
Db     89 LGYLANLTELDLSQNSFSDARDCDFHALPQLLSLHLEENQLTRLEDHSFAGLASLQELYL 148

Qy    112 RNNPIESIPSYAFNRIPSLRRLDLGE-------------LKRLSYISEGA---------- 148
           :| :  |   ||: : :| || |                |  |  :  |
Db    149 NHNQLYRIAPRAFSGLSNLLRLHLNSNLLRAIDSRWFEMLPNLEILMIGGNKVDAILDMN 208

Qy    149 FEGLSNLRYLNLAMCNLREIPN-------------------------LTPLIKLDELDL 182
          |  |:||| | ||  ||||| :                         |  :  |  |||
Db    209 FRPLANLRSLVLAGMNLREISDYALEGLQSLESLSFYDNQLARVPRRALEQVPGLKFLDL 268

Qy    183 SGNHLSAIRPGSFQGLMHLQKLWMIQSQ-------------------------IQVIER 216
          : | |  : || |  ::||::| :  :                          : |
Db    269 NKNPLQRVGPGDFANMLHLKELGLNNMEELVSIDKFALVNLPELTKLDITNNPRLSFIHP 328

Qy    217 NAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLS---WWIK 273
          || :|  :  : | :| |: |       | :|: : || ||  |:| | | :     ::
Db    329 RAFHHLPQMETLMLNNNALSALHQQTVESLPNLQEVGLHGNPIRCDCVIRWANATGTRVR 388

Qy    274 DMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEP---PADLNVTEGMAAELK 330
          : | :| |      ||:|:   : |:     | :   :: |   |  | |  | :  |
Db    389 FIEPQSTLCA----EPPDLQRLPVREVPFREMTDHCLPLISPRSFPPSLQVASGESMVLH 444

Qy    331 CRA-STSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVG 389
          |||  :    : |:|| |  :|      |  |  :|||    || :: |:|||:  | ||
Db    445 CRALAEPEPEIYWVTPAGLRLTPAHAGRRCRVYPEGTLELRRVTAEEAGLYTCVAQNLVG 504

Qy    390 NTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTP----VVDWETTNV 445
            | : :: |  |               ::| :|| :  :  |  |     :: |
Db    505 ADTKTVSVVVGRA-------------LLQPGRDEGQGLELRVQETHPYHILLSW----- 545

Qy    446 TTSLTPQSTRSTEKTFT 462
            :|| :| ||  |::
Db    546 ---VTPPNTVSTNLTWS 559


RESULT   2
```

```
CHAD_BOVIN
ID   CHAD_BOVIN     STANDARD;      PRT;   361 AA.
AC   Q27972;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Chondroadherin precursor (Cartilage leucine-rich protein) (38 kDa bone
DE   protein).
GN   CHAD.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Cartilage;
RX   MEDLINE=94342341; PubMed=8063792;
RA   Neame P.J., Sommarin Y., Boynton R.E., Heinegaard D.;
RT   "The structure of a 38-kDa leucine-rich protein (chondroadherin)
RT   isolated from bovine cartilage.";
RL   J. Biol. Chem. 269:21547-21554(1994).
RN   [2]
RP   SEQUENCE OF 25-55 AND 77-97.
RC   TISSUE=Bone;
RX   MEDLINE=95113864; PubMed=7814406;
RA   Hu B., Coulson L., Moyer B., Price P.A.;
RT   "Isolation and molecular cloning of a novel bone phosphoprotein
RT   related in sequence to the cystatin family of thiol protease
RT   inhibitors.";
RL   J. Biol. Chem. 270:431-436(1995).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U08018; AAA21330.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 10.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 5.
KW   Repeat; Signal.
FT   SIGNAL        1     24       OR 23 (IN SOME ISOFORM(S)).
FT   CHAIN        25    361       CHONDROADHERIN.
FT   CHAIN        25    352       CHONDROADHERIN, MINOR FORM.
FT   DOMAIN       79    317       10 X 24 AA LEUCINE-RICH TANDEM REPEATS.
FT   REPEAT       79    102       1.
FT   REPEAT      103    126       2.
FT   REPEAT      127    150       3.
FT   REPEAT      151    174       4.
FT   REPEAT      175    198       5.
FT   REPEAT      199    222       6.
FT   REPEAT      223    246       7.
FT   REPEAT      248    271       8.
FT   REPEAT      272    293       9.
FT   REPEAT      294    317       10.
FT   DISULFID    306    348
FT   DISULFID    308    328
FT   CONFLICT     25     25       C -> Y (IN REF. 2).
FT   CONFLICT     29     29       C -> W (IN REF. 2).
FT   CONFLICT     31     31       C -> H (IN REF. 2).
FT   CONFLICT     40     40       C -> L (IN REF. 2).
FT   CONFLICT     52     52       S -> R (IN REF. 2).
```

```
SQ   SEQUENCE   361 AA;  40884 MW;  DA79DC98AD3DD1F8 CRC64;

  Query Match           15.6%;  Score 394;  DB 1;  Length 361;
  Best Local Similarity   31.9%;  Pred. No. 4.8e-21;
  Matches  104;  Conservative   49;  Mismatches  139;  Indels   34;  Gaps    8;

Qy       3 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFK--------H 54
           ||  | | :    ||| :  |:::|  :|  |:|||| |   ::  |||:        |
Db      25 CPQNCHCHSDLQHVICDKVGLQKIPK-VSEKTKLLNLQRNNFPVLATNSFRAMPNLVSLH 83

Qy      55 LRHLEI----------------LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNG 98
           |:| :|                | || | || :  |||: |  |  | |  |::| :| |
Db      84 LQHCQIREVAAGAFRGLKQLIYLYLSHNDIRVLRAGAFDDLTELTYLYLDHNKVTELPRG 143

Qy      99 AFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYL 158
             |  |  | | || |  : | ||     || | | |   || :  || : : ||
Db     144 LLSPLVNLFILQLNNNKIRELRSGAFQGAKDLRWLYLSE-NSLSSLQPGALDDVENLAKF 202

Qy     159 NLAMCNLREIPN--LTPLIKLDELDLSGNHLSAIRPGSFQGL-MHLQKLWMIQSQIQVIE 215
           |    |   |:  |: |  ::|| || | | :|   :||    :|: ||:  : ::
Db     203 YLDRNQLSSYPSAALSKLRVVEELKLSHNPLKSIPDNAFQSFGRYLETLWLDNTNLEKFS 262

Qy     216 RNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDM 275
             ||  : :|  ::| :| |  || :   |   || : | :||| | | :  |  |::
Db     263 DGAFLGVTTLKHVHLENNRLHQLPSNF--PFDSLETLTLTNNPWKCTCQLRGLRRWLEAK 320

Qy     276 APSNTACCARCNTPPNLKGRYIGELD 301
             | ||   :|   :|::| : |
Db     321 TSRPDATCA---SPAKFRGQHIRDTD 343


RESULT   3
GPV_MOUSE
ID   GPV_MOUSE      STANDARD;      PRT;   567 AA.
AC   O08742;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Platelet glycoprotein V precursor (GPV) (CD42D).
GN   GP5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Liver;
RX   MEDLINE=97275136; PubMed=9129030;
RA   Ravanat C., Morales M., Azorsa D.O., Moog S., Schuhler S.,
RA   Grunert P., Loew D., van Dorsselaer A., Cazenave J.-P., Lanza F.;
RT   "Gene cloning of rat and mouse platelet glycoprotein V:
RT   identification of megakaryocyte-specific promoters and demonstration
RT   of functional thrombin cleavage.";
RL   Blood 89:3253-3262(1997).
CC   -!- FUNCTION: THE GPIB-V-IX COMPLEX FUNCTIONS AS THE VON WILLEBRAND
CC       FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT
CC       PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO
CC       INJURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A
CC       CRITICAL INITIATING EVENT IN HEMOSTASIS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: CONTAINS 14 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z69595; CAA93441.1; -.
DR   MGD; MGI:1096363; Gp5.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 13.
DR   Pfam; PF01463; LRRCT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00370; LRR; 1.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 10.
KW   Platelet; Transmembrane; Glycoprotein; Blood coagulation;
KW   Repeat; Leucine-rich repeat; Cell adhesion; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    567       PLATELET GLYCOPROTEIN V.
FT   DOMAIN       17    522       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    523    543       POTENTIAL.
FT   DOMAIN      544    567       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       73     96       LRR 1.
FT   REPEAT       97    120       LRR 2.
FT   REPEAT      122    144       LRR 3.
FT   REPEAT      145    168       LRR 4.
FT   REPEAT      170    192       LRR 5.
FT   REPEAT      194    216       LRR 6.
FT   REPEAT      217    240       LRR 7.
FT   REPEAT      241    264       LRR 8.
FT   REPEAT      266    288       LRR 9.
FT   REPEAT      289    312       LRR 10.
FT   REPEAT      314    337       LRR 11.
FT   REPEAT      338    361       LRR 12.
FT   REPEAT      362    385       LRR 13.
FT   REPEAT      386    409       LRR 14.
FT   CARBOHYD     51     51       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     67     67       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    243    243       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    298    298       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    312    312       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    385    385       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   567 AA;  63467 MW;  C48643AA73967A7D CRC64;

  Query Match           14.5%;  Score 367;  DB 1;  Length 567;
  Best Local Similarity   31.3%;  Pred. No. 7e-19;
  Matches  103;  Conservative   53;  Mismatches  133;  Indels   40;  Gaps    9;

Qy      20 RKNLREVPDGI---STNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAF 76
           | ||  :| |:          | |: ||:  :      :|  |  |:| |||:|::  |||
Db     179 RNNLTHLPKGLLGAQVKLEKLLLYSNQLTSVDSGLLSNLGALTELRLERNHLRSVAPGAF 238

Qy      77 NGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLG 136
           : | ||:|| |  | | ::|   |:::| :  | |  ||:| :|   |  :  || | |
Db     239 DRLGNLSSLTLSGNLLESLPPALFLHVSSVSRLTLFENPLEELPDVLFGEMAGLRELWLN 298

Qy     137 ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIK-----LDELDLSGNH---LS 188
             || :   ||  || |: | |        | |: | :     | || : | |   |:
Db     299 G-THLSTLPAAAFRNLSGLQTLGLT-----RNPRLSALPRGVFQGLRELRVLGLHTNALA 352

Qy     189 AIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHH 248
           :|  : :|| ||::: :  :::: : |  | || ||  : | || |  || |:|  |
Db     353 ELRDDALRGLGHLRQVSLRHNRLRALPRTLFRNLSSLESVQLEHNQLETLPGDVFAALPQ 412

Qy     249 LERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCY 308
           | :: | |||| |:| :     |::                |:: ||   |  |    |
Db     413 LTQVLLGHNPWLCDCGLWRFLQWLRH-------------HPDILGR--DEPPQ----CR 452

Qy     309 APVIVEPPADLNVTEGMAAELKCRASTSL 337
             |   || | |:  | :  :  |    ||
Db     453 GP---EPRASLSFWELLQGDPWCPDPRSL 478


RESULT   4
```

```
GPV_RAT
ID   GPV_RAT        STANDARD;      PRT;   567 AA.
AC   O08770;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Platelet glycoprotein V precursor (GPV) (CD42D).
GN   GP5.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WISTAR; TISSUE=Liver;
RX   MEDLINE=97275136; PubMed=9129030;
RA   Ravanat C., Morales M., Azorsa D.O., Moog S., Schuhler S.,
RA   Grunert P., Loew D., van Dorsselaer A., Cazenave J.-P., Lanza F.;
RT   "Gene cloning of rat and mouse platelet glycoprotein V:
RT   identification of megakaryocyte-specific promoters and demonstration
RT   of functional thrombin cleavage.";
RL   Blood 89:3253-3262(1997).
CC   -!- FUNCTION: THE GPIB-V-IX COMPLEX FUNCTIONS AS THE VON WILLEBRAND
CC       FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT
CC       PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO
CC       INJURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A
CC       CRITICAL INITIATING EVENT IN HEMOSTASIS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: CONTAINS 14 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z69594; CAA93440.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 14.
DR   Pfam; PF01463; LRRCT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00370; LRR; 2.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 10.
KW   Platelet; Transmembrane; Glycoprotein; Blood coagulation;
KW   Repeat; Leucine-rich repeat; Cell adhesion; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    567       PLATELET GLYCOPROTEIN V.
FT   DOMAIN       17    522       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    523    543       POTENTIAL.
FT   DOMAIN      544    567       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       73     96       LRR 1.
FT   REPEAT       97    120       LRR 2.
FT   REPEAT      122    144       LRR 3.
FT   REPEAT      145    168       LRR 4.
FT   REPEAT      169    192       LRR 5.
FT   REPEAT      194    216       LRR 6.
FT   REPEAT      217    240       LRR 7.
FT   REPEAT      241    264       LRR 8.
FT   REPEAT      266    288       LRR 9.
FT   REPEAT      289    312       LRR 10.
FT   REPEAT      314    337       LRR 11.
FT   REPEAT      338    361       LRR 12.
FT   REPEAT      362    385       LRR 13.
FT   REPEAT      387    409       LRR 14.
FT   CARBOHYD     51     51       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    243    243       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    298    298       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    312    312       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    385    385       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    498    498       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   567 AA;  63344 MW;  CA10708E0D03707F CRC64;


  Query Match             14.1%;  Score 356;  DB 1;  Length 567;
  Best Local Similarity   29.9%;  Pred. No. 4.3e-18;
  Matches  110;  Conservative   53;  Mismatches  163;  Indels   42;  Gaps    9;

Qy       6 VCSCSNQFS------------KVICV-RKNLREVPDGI---STNTRLLNLHENQIQIIK 48
           :|   || |              ||: : | ||  :| |:         | |: |::  :
Db     151 LCLNQNQLSFLPANLFSSLGKLKVLDLSRNNLTHLPQGLLGAQIKLEKLLLYSNRLMSLD 210

Qy      49 VNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKE 108
             :|  |  |:| |||:|:|  |||: | ||:|| |  | | ::|   |:::| |
Db     211 SGLLANLGALTELRLERNHLRSIAPGAFDSLGNLSTLTLSGNLLESLPPALFLHVSWLTR 270

Qy     109 LWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREI 168
           | |  ||:| :|   |  :  || | |     |  :   ||  || |: | |    | :
Db     271 LTLFENPLEELPEVLFGEMAGLRELWLNG-THLRTLPAAAFRNLSGLQTLGLTRNPL--L 327

Qy     169 PNLTP-----LIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQ 223
             | |     | :|  | :  | |  :   : :||  |::: :  :::: : |  | ||
Db     328 SALPPGMFHGLTELRVLAVHTNALEELPEDALRGLGRLRQVSLRHNRLRALPRTLFRNLS 387

Qy     224 SLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWI--------KDM 275
           ||| : | || |  || |:|  |  | |: | |||| |:| :     |:        :|
Db     388 SLVTVQLEHNQLKTLPGDVFAALPQLTRVLLGHNPWLCDCGLWPFLQWLRHHLELLGRDE 447

Qy     276 APSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRAST 335
           |       :|| | :       || |    | |     | |      : |    :
Db     448 PP-------QCNGPESRASLTFWELLQGDQWC--PSSRGLPPDPPTENALKAPDPTQRPN 498

Qy     336 SLTSVSWI 343
           |  | :|:
Db     499 SSQSWAWV 506


RESULT   5
SLIT_DROME
ID   SLIT_DROME     STANDARD;      PRT;  1480 AA.
AC   P24014;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Slit protein precursor.
GN   SLI.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91099665; PubMed=2176636;
RA   Rothberg J.M., Jacobs J.R., Goodman C.S., Artavanis-Tsakonas S.;
RT   "Slit: an extracellular protein necessary for development of midline
RT   glia and commissural axon pathways contains both EGF and LRR
RT   domains.";
RL   Genes Dev. 4:2169-2187(1990).
CC   -!- FUNCTION: NECESSARY FOR DEVELOPMENT OF MIDLINE GLIA AND
CC       COMMISSURAL AXON PATHWAYS. SLIT MAY INTERACT WITH EXTRACELLULAR
CC       MATRIX MOLECULES.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXCRETED BY THE MIDLINE GLIA CELLS AND
CC       EVENTUALLY DISTRIBUTED ALONG THE AXONS.
CC   -!- SIMILARITY: CONTAINS 7 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 24 LEUCINE-RICH REPEATS (LRR).
```

```
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   -!- SIMILARITY: CONTAINS 1 LAMININ G-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X53959; CAA37910.1; -.
DR   PIR; A36665; A36665.
DR   HSSP; P00743; 1CCF.
DR   FlyBase; FBgn0003425; sli.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000359; Cys_knot.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00007; Cys_knot; 1.
DR   Pfam; PF00008; EGF; 7.
DR   Pfam; PF00054; laminin_G; 1.
DR   Pfam; PF00560; LRR; 17.
DR   Pfam; PF01463; LRRCT; 4.
DR   Pfam; PF01462; LRRNT; 4.
DR   SMART; SM00041; CT; 1.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00001; EGF_like; 5.
DR   SMART; SM00370; LRR; 4.
DR   SMART; SM00082; LRRCT; 4.
DR   SMART; SM00013; LRRNT; 4.
DR   SMART; SM00369; LRR_TYP; 9.
DR   SMART; SM00282; LamG; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS00022; EGF_1; 7.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 1.
KW   Neurogenesis; Glycoprotein; Signal; Alternative splicing;
KW   EGF-like domain; Repeat; Leucine-rich repeat.
FT   SIGNAL       1   36
FT   CHAIN       37  1480       SLIT PROTEIN.
FT   REPEAT      99   122        LRR 1.
FT   REPEAT     123   146        LRR 2.
FT   REPEAT     148   170        LRR 3.
FT   REPEAT     171   194        LRR 4.
FT   REPEAT     195   218        LRR 5.
FT   REPEAT     220   246        LRR 6.
FT   REPEAT     321   344        LRR 7.
FT   REPEAT     345   368        LRR 8.
FT   REPEAT     369   392        LRR 9.
FT   REPEAT     394   416        LRR 10.
FT   REPEAT     417   440        LRR 11.
FT   REPEAT     522   545        LRR 12.
FT   REPEAT     546   569        LRR 13.
FT   REPEAT     570   593        LRR 14.
FT   REPEAT     595   617        LRR 15.
FT   REPEAT     618   641        LRR 16.
FT   REPEAT     643   666        LRR 17.
FT   REPEAT     678   701        LRR 18.
FT   REPEAT     720   743        LRR 19.
FT   REPEAT     745   764        LRR 20.
FT   REPEAT     765   788        LRR 21.
FT   REPEAT     790   812        LRR 22.
FT   REPEAT     813   836        LRR 23.
FT   REPEAT     838   861        LRR 24.
FT   DOMAIN     907   944        EGF-LIKE 1.
FT   DOMAIN     946   983        EGF-LIKE 2.
FT   DOMAIN     985  1022        EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1024  1062        EGF-LIKE 4.
FT   DOMAIN    1064  1100        EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1111  1149        EGF-LIKE 6.
FT   DOMAIN    1152  1325        LAMININ G-LIKE.
FT   DOMAIN    1353  1392        EGF-LIKE 7.
FT   DOMAIN    1409  1480        CTCK.
FT   CARBOHYD   111   111        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   207   207        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   357   357        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   435   435        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   783   783        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   788   788        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   958   958        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   998   998        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1060  1060        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1159  1159        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1175  1175        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1243  1243        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1292  1292        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID   911   922        BY SIMILARITY.
FT   DISULFID   916   932        BY SIMILARITY.
FT   DISULFID   934   943        BY SIMILARITY.
FT   DISULFID   950   961        BY SIMILARITY.
FT   DISULFID   955   971        BY SIMILARITY.
FT   DISULFID   973   982        BY SIMILARITY.
FT   DISULFID   989  1001        BY SIMILARITY.
FT   DISULFID   995  1010        BY SIMILARITY.
FT   DISULFID  1012  1021        BY SIMILARITY.
FT   DISULFID  1028  1041        BY SIMILARITY.
FT   DISULFID  1035  1050        BY SIMILARITY.
FT   DISULFID  1052  1061        BY SIMILARITY.
FT   DISULFID  1068  1079        BY SIMILARITY.
FT   DISULFID  1073  1088        BY SIMILARITY.
FT   DISULFID  1090  1099        BY SIMILARITY.
FT   DISULFID  1115  1125        BY SIMILARITY.
FT   DISULFID  1120  1137        BY SIMILARITY.
FT   DISULFID  1139  1148        BY SIMILARITY.
FT   DISULFID  1357  1368        BY SIMILARITY.
FT   DISULFID  1362  1380        BY SIMILARITY.
FT   DISULFID  1382  1391        BY SIMILARITY.
FT   DISULFID  1409  1443        BY SIMILARITY.
FT   DISULFID  1423  1457        BY SIMILARITY.
FT   DISULFID  1434  1473        BY SIMILARITY.
FT   DISULFID  1438  1475        BY SIMILARITY.
FT   DISULFID  1442  1479        BY SIMILARITY.
FT   VARSPLIC  1394  1404        MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE  1480 AA;  165752 MW;  F9D5925FC170B1C3 CRC64;


  Query Match           13.9%;  Score 350;  DB 1;  Length 1480;
  Best Local Similarity  23.9%;  Pred. No. 3.9e-17;
  Matches  109;  Conservative   65;  Mismatches  152;  Indels  130;  Gaps    9;

Qy       2 TCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEIL 61
             :||  | |::    | |  |:|  ||  :  :|  : | :| |  :   ||   | |  :
Db     294 SCPHPCRCADGI--VDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKSFSSFRRLRRI 351

Qy      62 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 121
             || |:|  |   | :||  | || |: |::  :|:| |  |  |: | |  | |  |
Db     352 DLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLLNANEISCIRK 411

Qy     122 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM------CNLR--------- 166
             ||  : ||  | | :   :  :: | |: : ::: ::||       ||||
Db     412 DAFRDLHSLSLLSLYD-NNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRWLADYLHKN 470

Qy     167 ---------------------------------------------------------- 166

Db     471 PIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKLSGECRMDSDCPAMCHCEGTTV 530
```

```
Qy    167 --------EIPNLTPL--------------IKLD----------ELDLSGNHLSAIRPGS 194
                  |||   ||              |  |          :|:|  | |: | | :
Db    531 DCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKRNQLTGIEPNA 590

Qy    195 FQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 254
          |:|  |:|:| : :::|: |    |  |  |  :||  | :: :    |  |: |  ::|
Db    591 FEGASHIQELQLGENKIKEISNKMFLGLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNL 650

Qy    255 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVE 314
            ||:|||| : | :  ::  : :  |  |||  |  ::   | :|  : | | :
Db    651 ASNPFNCNCHLAWFAECVRKKSLNGGA--ARCGAPSKVRDVQIKDLPHSEFKCSS----- 703

Qy    315 PPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVM 350
                :||   :  |  | : |        |||:
Db    704 -----ENSEGCLGDGYCPPSCTCT--------GTVV 726
```

```
RESULT   6
GPV_HUMAN
ID   GPV_HUMAN      STANDARD;      PRT;   560 AA.
AC   P40197;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Platelet glycoprotein V precursor (GPV) (CD42D).
GN   GP5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=93391348; PubMed=7690959;
RA   Hickey M.J., Hagen F.S., Yagi M., Roth G.J.;
RT   "Human platelet glycoprotein V: characterization of the polypeptide
RT   and the related Ib-V-IX receptor system of adhesive, leucine-rich
RT   glycoproteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:8327-8331(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Platelet;
RX   MEDLINE=94012616; PubMed=8407908;
RA   Lanza F., Morales M., de la Salle C., Cazenave J.-P., Clemetson K.J.,
RA   Shimomura T., Phillips D.R.;
RT   "Cloning and characterization of the gene encoding the human platelet
RT   glycoprotein V. A member of the leucine-rich glycoprotein family
RT   cleaved during thrombin-induced platelet activation.";
RL   J. Biol. Chem. 268:20801-20807(1993).
RN   [3]
RP   PARTIAL SEQUENCE.
RC   TISSUE=Platelet;
RX   MEDLINE=90275263; PubMed=2350580;
RA   Shimomura T., Fujimura K., Maehama S., Takemoto M., Oda K.,
RA   Fujimoto T., Oyama R., Suzuki M., Icihara-Tanaka K., Titani K.,
RA   Kuramoto A.;
RT   "Rapid purification and characterization of human platelet
RT   glycoprotein V: the amino acid sequence contains leucine-rich
RT   repetitive modules as in glycoprotein Ib.";
RL   Blood 75:2349-2356(1990).
RN   [4]
RP   PARTIAL SEQUENCE.
RC   TISSUE=Platelet;
RX   MEDLINE=90321220; PubMed=2372284;
RA   Roth G.J., Church T.A., McMullen B.A., Williams S.A.;
RT   "Human platelet glycoprotein V: a surface leucine-rich glycoprotein
RT   related to adhesion.";
RL   Biochem. Biophys. Res. Commun. 170:153-161(1990).
CC   -!- FUNCTION: THE GPIB-V-IX COMPLEX FUNCTIONS AS THE VON WILLEBRAND
CC       FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT
CC       PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO
CC       INJURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A
CC       CRITICAL INITIATING EVENT IN HEMOSTASIS.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: PLATELETS AND MEGAKARYOCYTES.
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   -!- SIMILARITY: CONTAINS 14 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L11238; AAA03069.1; -.
DR   EMBL; Z23091; CAA80637.1; -.
DR   HSSP; P16473; 1XUM.
DR   MIM; 173511; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 14.
DR   Pfam; PF01463; LRRCT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00370; LRR; 1.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 9.
KW   Platelet; Transmembrane; Glycoprotein; Blood coagulation;
KW   Repeat; Leucine-rich repeat; Cell adhesion; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    560       PLATELET GLYCOPROTEIN V.
FT   DOMAIN       17    523       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    524    544       POTENTIAL.
FT   DOMAIN      545    560       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       73     96       LRR 1.
FT   REPEAT       97    120       LRR 2.
FT   REPEAT      122    144       LRR 3.
FT   REPEAT      145    168       LRR 4.
FT   REPEAT      169    192       LRR 5.
FT   REPEAT      194    216       LRR 6.
FT   REPEAT      217    240       LRR 7.
FT   REPEAT      241    264       LRR 8.
FT   REPEAT      266    288       LRR 9.
FT   REPEAT      289    312       LRR 10.
FT   REPEAT      314    337       LRR 11.
FT   REPEAT      338    361       LRR 12.
FT   REPEAT      362    385       LRR 13.
FT   REPEAT      386    409       LRR 14.
FT   CARBOHYD     51     51       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    243    243       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    267    267       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    298    298       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    312    312       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    385    385       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    499    499       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     73     74       MT -> TK (IN REF. 2).
FT   CONFLICT    109    109       K -> T (IN REF. 2).
FT   CONFLICT    130    130       D -> W (IN REF. 3).
FT   CONFLICT    136    138       GID -> PGG (IN REF. 3).
FT   CONFLICT    209    209       L -> I (IN REF. 2).
FT   CONFLICT    267    267       N -> H (IN REF. 3).
FT   CONFLICT    327    327       L -> I (IN REF. 2).
FT   CONFLICT    478    478       P -> G (IN REF. 2).
FT   CONFLICT    509    509       P -> D (IN REF. 2).
SQ   SEQUENCE   560 AA;  60959 MW;  B1CDB04AF8AF7115 CRC64;
```

```
  Query Match             13.1%;  Score 329.5;  DB 1;  Length 560;
  Best Local Similarity   29.7%;  Pred. No. 3.2e-16;
```

```
  Matches   83;  Conservative   43;  Mismatches  104;  Indels   49;  Gaps    3;

Qy      22 NLREVPDG-----------------ISTNTRLLN---------LHENQIQIIKVNSFKH 54
           ||  :| |                  :| :: |||         | | |: |   :|
Db     181 NLTHLPKGLLGAQAKLERLLLHSNRLVSLDSGLLNSLGALTELQFHRNHIRSIAPGAFDR 240

Qy      55 LRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNN 114
           | :|  | |||||:  :    |    ||  | ||:| |  :|   |  :  |:|||| 
Db     241 LPNLSSLTLSRNHLAFLPSALFLHSHNLTLLTLFENPLAELPGVLFGEMGGLQELWLNRT 300

Qy     115 PIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPL 174
            : ::|: ||  :  || | :    ||| : :|||:||  |:
Db     301 QLRTLPAAAFRNLSRLRYLGVTLSPRLSALPQGAFQGLGELQV----------------- 343

Qy     175 IKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNN 234
               | |  | |:|:  |  :|| |::: : ::::: : |  | || ||  : | ||
Db     344 -----LALHSNGLTALPDGLLRGLGKLRQVSLRRNRLRALPRALFRNLSSLESVQLDHNQ 398

Qy     235 LTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIK 273
           |  || |:|  |  |  : | || | |:| :     |::
Db     399 LETLPGDVFGALPRLTEVLLGHNSWRCDCGLGPFLGWLR 437


RESULT  7
ALS_PAPHA
ID   ALS_PAPHA      STANDARD;      PRT;   605 AA.
AC   O02833;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Insulin-like growth factor binding protein complex acid labile chain
DE   precursor (ALS).
GN   IGFALS OR ALS.
OS   Papio hamadryas (Hamadryas baboon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Papio.
OX   NCBI_TaxID=9557;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=97040714; PubMed=8886027;
RA   Delhanty P., Baxter R.C.;
RT   "The cloning and expression of the baboon acid-labile subunit of the
RT   insulin-like growth factor binding protein complex.";
RL   Biochem. Biophys. Res. Commun. 227:897-902(1996).
CC   -!- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS THAT RESULT
CC       IN PROTEIN COMPLEXES, RECEPTOR-LIGAND BINDING OR CELL ADHESION.
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 kDa WITH
CC       IGF-I OR IGF-II AND IGFBP-3 (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- SIMILARITY: CONTAINS 20 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S83462; -; NOT_ANNOTATED_CDS.
DR   HSSP; P23945; 1XUN.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 19.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00370; LRR; 2.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 11.
KW   Glycoprotein; Leucine-rich repeat; Repeat; Signal.
FT   SIGNAL        1     27       BY SIMILARITY.
FT   CHAIN        28    605       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                PROTEIN COMPLEX ACID LABILE CHAIN.
FT   REPEAT       53     73       LRR 1.
FT   REPEAT       74     96       LRR 2.
FT   REPEAT       98    120       LRR 3.
FT   REPEAT      121    144       LRR 4.
FT   REPEAT      146    168       LRR 5.
FT   REPEAT      169    192       LRR 6.
FT   REPEAT      194    216       LRR 7.
FT   REPEAT      217    240       LRR 8.
FT   REPEAT      242    264       LRR 9.
FT   REPEAT      265    288       LRR 10.
FT   REPEAT      289    312       LRR 11.
FT   REPEAT      313    336       LRR 12.
FT   REPEAT      338    360       LRR 13.
FT   REPEAT      361    384       LRR 14.
FT   REPEAT      386    408       LRR 15.
FT   REPEAT      409    432       LRR 16.
FT   REPEAT      433    456       LRR 17.
FT   REPEAT      458    480       LRR 18.
FT   REPEAT      482    504       LRR 19.
FT   REPEAT      505    530       LRR 20.
FT   CARBOHYD     64     64       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     85     85       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     96     96       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    368    368       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    515    515       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    580    580       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   605 AA;  66110 MW;  9D71994625F23652 CRC64;


  Query Match             13.1%;  Score 329.5;  DB 1;  Length 605;
  Best Local Similarity   21.0%;  Pred. No. 3.5e-16;
  Matches  119;  Conservative   61;  Mismatches  118;  Indels  269;  Gaps   12;

Qy       3 CPSVCSCS-----NQFSKVICVRKNLREVPDGISTNTRL---------------------- 36
             ||: |:||     |: | | |  :||  :||||   |:
Db      41 CPATCACSYDDEVNELS-VFCSSRNLTRLPDGIPGGTQALWLDSNNLSSIPPAAFRNLSS 99

Qy      37 --------------------------LNLHENQIQIIKVNSFKHLRHLEILQLSRNHIR 69
                                       |:|  ||:: : | :| :   | :| || | :
Db     100 LAFLNLQGGQLGSLEPQALLGLENLCHLHLERNQLRSLAVGTFAYTPALALLGLSNNRLS 159

Qy      70 TIEIGAFNGLANLNTLELFDNRLTTIPNGAF----------------VY--------LSK 105
              :| | | || ||  | |  | |  :|: ||                |         |::
Db     160 RLEDGLFEGLGNLWDLNLGWNSLAVLPDAAFRGLGGLRELVLAGNRLAYLQPALFSGLAE 219

Qy     106 LKELWLRNNPIESIPSYAFNRIPSLRRL------------------------------ 133
           |:|| |  | : :| :  | ::| |::|
Db     220 LRELDLSRNALRAIKANVFAQLPRLQKLYLDRNLIAAVAPGAFLGLKALRWLDLSHNRVA 279

Qy     134 ------------------------------------------------------DLGE 137
                                                                   ||:
Db     280 GLLEDTFPGLLGLRVLRLSHNAIASLRPRTFEDLHFLEELQLGHNRIRQLAERSFEGLGQ 339

Qy     138 LK-------RLSYISEGAFEGLSNLRYLNLAMCNLREIPN--LTPLIKLDELDLSGNHLS 188
           |:       :|  :  ||| ||:|:  :||:   || :|      | ||  | | |: |
Db     340 LEVLTLDHNQLQEVKVGAFLGLTNVAVMNLSGNCLRNLPEQVFRGLGKLHSLHLEGSCLG 399

Qy     189 AIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPL-- 246
               ||| :| ||  |::|::  : :  ||  :   |  |:|::|  | || ||| ||  |
Db     400 RIRPHTFAGLSGLRRLFLKDNGLVGIEEQSLWGLAELLELDLTSNQLTHLPHQLFQGLGK 459

Qy     247 -------------------------------HH-------------------------- 248
                                           |:
Db     460 LEYLLLSHNRLAELPADALGPLQRAFWLDVSHNRLEALPGSLLASLGRLRYLNLRNNSLR 519

Qy     249 --------LERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCAR--------------- 285
```

```
                |||: |  |||:|:| :  |    :| |  | :   |
Db    520 TFTPQPPGLERLWLEGNPWDCSCPLKAL----RDFALQNPSAVPRFVQAICEGDDCQPPV 575

Qy    286 -------CNTPPNLKGRYIGELDQNYF 305
                | :|| : |  : :| : :|
Db    576 YTYNNITCASPPEVAGLDLRDLGEAHF 602


RESULT   8
ALS_HUMAN
ID   ALS_HUMAN      STANDARD;      PRT;   605 AA.
AC   P35858;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Insulin-like growth factor binding protein complex acid labile chain
DE   precursor (ALS).
GN   IGFALS OR ALS.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Liver;
RX   MEDLINE=92357025; PubMed=1379671;
RA   Leong S.R., Baxter R.C., Camerato T., Dai J., Wood W.I.;
RT   "Structure and functional expression of the acid-labile subunit of
RT   the insulin-like growth factor-binding protein complex.";
RL   Mol. Endocrinol. 6:870-876(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Frankland J.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 28-35.
RX   MEDLINE=89308584; PubMed=2473065;
RA   Baxter R.C., Martin J.L., Beniac V.A.;
RT   "High molecular weight insulin-like growth factor binding protein
RT   complex. Purification and properties of the acid-labile subunit from
RT   human serum.";
RL   J. Biol. Chem. 264:11843-11848(1989).
CC   -!- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS THAT RESULT
CC       IN PROTEIN COMPLEXES, RECEPTOR-LIGAND BINDING OR CELL ADHESION.
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 kDa WITH
CC       IGF-I OR IGF-II AND IGFBP-3.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: CONTAINS 20 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M86826; AAA36047.1; -.
DR   EMBL; AL031724; CAC36078.1; -.
DR   PIR; A41915; A41915.
DR   HSSP; P23945; 1XUN.
DR   MIM; 601489; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 19.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00370; LRR; 2.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 11.
KW   Glycoprotein; Leucine-rich repeat; Repeat; Signal.
FT   SIGNAL        1     27
FT   CHAIN        28    605       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                 PROTEIN COMPLEX ACID LABILE CHAIN.
FT   REPEAT        53     73       LRR 1.
FT   REPEAT        74     96       LRR 2.
FT   REPEAT        98    120       LRR 3.
FT   REPEAT       121    144       LRR 4.
FT   REPEAT       145    168       LRR 5.
FT   REPEAT       170    192       LRR 6.
FT   REPEAT       193    216       LRR 7.
FT   REPEAT       217    240       LRR 8.
FT   REPEAT       242    264       LRR 9.
FT   REPEAT       265    288       LRR 10.
FT   REPEAT       289    312       LRR 11.
FT   REPEAT       313    336       LRR 12.
FT   REPEAT       337    360       LRR 13.
FT   REPEAT       361    384       LRR 14.
FT   REPEAT       386    408       LRR 15.
FT   REPEAT       409    432       LRR 16.
FT   REPEAT       433    456       LRR 17.
FT   REPEAT       458    480       LRR 18.
FT   REPEAT       482    504       LRR 19.
FT   REPEAT       505    530       LRR 20.
FT   CARBOHYD      64     64       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      85     85       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      96     96       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     368    368       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     515    515       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     580    580       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   605 AA;  66034 MW;  F6562A23CBE918F6 CRC64;


  Query Match             12.5%;  Score 314.5;  DB 1;  Length 605;
  Best Local Similarity   31.1%;  Pred. No. 4.1e-15;
  Matches   89;  Conservative   47;  Mismatches  117;  Indels   33;  Gaps    5;

Qy      3 CPSVCSCS-----NQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRH 57
            ||: | ||     :: | | |  :||  :|||:   |: | |  | :  :   :|::|
Db     41 CPAACVCSYDDDADELS-VFCSSRNLTRLPDGVPGGTQALWLDGNNLSSVPPAAFQNLSS 99

Qy     58 LEI-----------------------LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLT 93
          |                         | | || :|:: :| |     | :| | :|||:
Db    100 LGFLNLQGGQLGSLEPQALLGLENLCHLHLERNQLRSLALGTFAHTPALASLGLSNNRLS 159

Qy     94 TIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLS 153
          : :| |  |  | :| |  | :  :|  ||  : ||| | |    ||:|:    | ||:
Db    160 RLEDGLFEGLGSLWDLNLGWNSLAVLPDAAFRGLGSLRELVLAG-NRLAYLQPALFSGLA 218

Qy    154 NLRYLNLAMCNLREIPN--LTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQI 211
           || |:|:   || |       | :| :| |  | ::|: ||:| ||  |: | :  :::
Db    219 ELRELDLSRNALRAIKANVFVQLPRLQKLYLDRNLIAAVAPGAFLGLKALRWLDLSHNRV 278

Qy    212 QVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHN 257
            :  : |  |  |  : |:|| :  |    |  || || : | ||
Db    279 AGLLEDTFPGLLGLRVLRLSHNAIASLRPRTFKDLHFLEELQLGHN 324


RESULT   9
ALS_RAT
ID   ALS_RAT        STANDARD;      PRT;   603 AA.
AC   P35859;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Insulin-like growth factor binding protein complex acid labile chain
DE   precursor (ALS).
GN   IGFALS OR ALS.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=93038676; PubMed=1384485;
RA   Dai J., Baxter R.C.;
RT   "Molecular cloning of the acid-labile subunit of the rat insulin-like
RT   growth factor binding protein complex.";
RL   Biochem. Biophys. Res. Commun. 188:304-309(1992).
RN   [2]
RP   SEQUENCE OF 24-44, AND CHARACTERIZATION.
RC   STRAIN=WISTAR; TISSUE=Serum;
RX   MEDLINE=94130835; PubMed=7507839;
RA   Baxter R.C., Dai J.;
RT   "Purification and characterization of the acid-labile subunit of rat
RT   serum insulin-like growth factor binding protein complex.";
RL   Endocrinology 134:848-852(1994).
CC   -!- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN REGULATING THE ACCESS OF
CC       CIRCULATING IGFS TO THE TISSUES.
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 kDa WITH
CC       IGF-I OR IGF-II AND IGFBP-3.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: BRAIN, KIDNEY, LUNG, HEART, SPLEEN, MUSCLE
CC       AND LIVER.
CC   -!- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S46785; AAB23770.2; -.
DR   PIR; JC1282; JC1282.
DR   HSSP; P23945; 1XUN.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 19.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00370; LRR; 5.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 9.
KW   Glycoprotein; Leucine-rich repeat; Repeat; Signal.
FT   SIGNAL       1     23
FT   CHAIN       24    603       INSULIN-LIKE GROWTH FACTOR BINDING
FT                               PROTEIN COMPLEX ACID LABILE CHAIN.
FT   REPEAT       52     73       LRR 1.
FT   REPEAT       74     96       LRR 2.
FT   REPEAT       98    120       LRR 3.
FT   REPEAT      121    144       LRR 4.
FT   REPEAT      146    168       LRR 5.
FT   REPEAT      169    192       LRR 6.
FT   REPEAT      194    216       LRR 7.
FT   REPEAT      217    240       LRR 8.
FT   REPEAT      242    264       LRR 9.
FT   REPEAT      266    288       LRR 10.
FT   REPEAT      289    312       LRR 11.
FT   REPEAT      313    336       LRR 12.
FT   REPEAT      337    360       LRR 13.
FT   REPEAT      361    384       LRR 14.
FT   REPEAT      386    408       LRR 15.
FT   REPEAT      409    432       LRR 16.
FT   REPEAT      433    456       LRR 17.
FT   REPEAT      458    478       LRR 18.
FT   REPEAT      479    504       LRR 19.
FT   REPEAT      506    529       LRR 20.
FT   REPEAT      543    566       LRR 21.
FT   CARBOHYD     64     64       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     85     85       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     96     96       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    368    368       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    515    515       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    586    586       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   603 AA;  66811 MW;  DCD7637D94A5037C CRC64;


  Query Match             12.2%;  Score 309;  DB 1;  Length 603;
  Best Local Similarity   30.3%;  Pred. No. 1e-14;
  Matches   96;  Conservative   43;  Mismatches  126;  Indels   52;  Gaps    8;

Qy      36 LLNLH-----ENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDN 90
           || ||      | |  ::  :|| |  || |||  | || :    | ||  |  | | ||
Db     289 LLGLHVLRLAHNAIASLRPRTFKDLHFLEELQLGHNRIRQLGERTFEGLGQLEVLTLNDN 348

Qy      91 RLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFE 150
           ::| :  |||  |  :  : |  | : |:|   |  :  |  | | |   | ::   |
Db     349 QITEVRVGAFSGLFNVAVMNLSGNCLRSLPERVFQGLDKLHSLHL-EHSCLGHVRLHTFA 407

Qy     151 GLSNLRYLNLAMCNLREI--PNLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQ 208
           ||| || | |   ::  |   :|  | :| |||: | |: :    ||| ||: | :
Db     408 GLSGLRRLFLRDNSISSIEEQSLAGLSELLELDLTTNRLTHLPRQLFQGLGHLEYLLLSY 467

Qy     209 SQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDL------------------FTPLHHL 249
           :|:  :     ||    ::::||:|  |  |                   |:|  |
Db     468 NQLTTLSAEVLGPLQRAFWLDISHNHLETLAEGLFSSLGRVRYLSLRNNSLQTFSPQPGL 527

Qy     250 ERIHLHHNPWNCNCDILWLSWWIKDMAPSN---------TAC----C--------ARCNT 288
           ||: |  |||:|:|: :  |    :| | |          | |    |          |
Db     528 ERLWLDANPWDCSCPLKAL----RDFALQNPGVVPRFVQTVCEGDDCQPVYTYNNITCAG 583

Qy     289 PPNLKGRYIGELDQNYF 305
           | |: |  : :: : :|
Db     584 PANVSGLDLRDVSETHF 600


RESULT  10
ALS_MOUSE
ID   ALS_MOUSE      STANDARD;      PRT;   603 AA.
AC   P70389;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Insulin-like growth factor binding protein complex acid labile chain
DE   precursor (ALS).
GN   IGFALS OR ALS OR ALBS.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SV;
RX   MEDLINE=96413591; PubMed=8816745;
RA   Boisclair Y.R., Seto D., Hsieh S., Hurst K.R., Ooi G.T.;
RT   "Organization and chromosomal localization of the gene encoding the
RT   mouse acid labile subunit of the insulin-like growth factor binding
RT   complex.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:10028-10033(1996).
CC   -!- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN REGULATING THE ACCESS OF
CC       CIRCULATING IGFS TO THE TISSUES.
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 kDa WITH
CC       IGF-I OR IGF-II AND IGFBP-3 (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U66900; AAB17270.1; -.
DR   MGD; MGI:107973; Igfals.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 19.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00370; LRR; 4.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 9.
KW   Glycoprotein; Leucine-rich repeat; Repeat; Signal.
FT   SIGNAL        1     23       BY SIMILARITY.
FT   CHAIN        24    603       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                PROTEIN COMPLEX ACID LABILE CHAIN.
FT   REPEAT       52     73       LRR 1.
FT   REPEAT       74     96       LRR 2.
FT   REPEAT       98    120       LRR 3.
FT   REPEAT      121    144       LRR 4.
FT   REPEAT      146    168       LRR 5.
FT   REPEAT      169    192       LRR 6.
FT   REPEAT      194    216       LRR 7.
FT   REPEAT      217    240       LRR 8.
FT   REPEAT      242    264       LRR 9.
FT   REPEAT      265    288       LRR 10.
FT   REPEAT      289    312       LRR 11.
FT   REPEAT      313    336       LRR 12.
FT   REPEAT      338    360       LRR 13.
FT   REPEAT      361    384       LRR 14.
FT   REPEAT      386    408       LRR 15.
FT   REPEAT      409    432       LRR 16.
FT   REPEAT      433    456       LRR 17.
FT   REPEAT      458    480       LRR 18.
FT   REPEAT      482    504       LRR 19.
FT   REPEAT      505    529       LRR 20.
FT   REPEAT      543    566       LRR 21.
FT   CARBOHYD     64     64       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     85     85       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     96     96       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    368    368       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    515    515       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    586    586       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   603 AA;  66959 MW;  4A6223AEC7026BCC CRC64;


  Query Match             12.0%;  Score 302.5;  DB 1;  Length 603;
  Best Local Similarity   31.5%;  Pred. No. 2.9e-14;
  Matches   93;  Conservative   41;  Mismatches  110;  Indels   51;  Gaps    7;

Qy       3 CPSVCSCS-----NQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRH 57
           ||  |:||     :: | | |  :|| ::|||  :|| | |  | :  |   :|::|
Db      41 CPVTCTCSYDDYTDELS-VFCSSRNLTQLPDGIPVSTRALWLDGNNLSSIPSAAFQNLSS 99

Qy      58 LEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIE 117
           |: | |  : :|::|  |  || ||  | |  | | ::  | | :   |  | | || :
Db     100 LDFLNLQGSWLRSLEPQALLGLQNLYHLHLERNLLRSLAAGLFRHTPSLASLSLGNNLLG 159

Qy     118 SIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLA----------MCNLR 166
           :    |  :  |  |:||    |  : :  |:|| ||  | ||           :| |
Db     160 RLEEGLFRGLSHLWDLNLG-WNSLVVLPDTVFQGLGNLHELVLAGNKLTYLQPALLCGLG 218

Qy     167 EIPNLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAF------- 219
           |         | ||||| | | :::   |  |  ||||:: :: |  :   ||
Db     219 E---------LRELDLSRNALRSVKANVFIHLPRLQKLYLDRNLITAVAPRAFLGMKALR 269

Qy     220 ------DNLQSLVE----------INLAHNNLTLLPHDLFTPLHHLERIHLHHN 257
                 : :  |:|           : |||| :| |    |  || || : | ||
Db     270 WLDLSHNRVAGLLEDTFPGLLGLHVLRLAHNAITSLRPRTFKDLHFLEELQLGHN 324


RESULT  11
PGS2_HUMAN
ID   PGS2_HUMAN     STANDARD;      PRT;   359 AA.
AC   P07585; Q9Y5N9; Q9Y5N8; Q9P0Z0; Q9P0Z1;
DT   01-APR-1988 (Rel. 07, Created)
DT   01-APR-1988 (Rel. 07, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Bone proteoglycan II precursor (PG-S2) (Decorin) (PG40).
GN   DCN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87017013; PubMed=3484330;
RA   Krusius T., Ruoslahti E.;
RT   "Primary structure of an extracellular matrix proteoglycan core
RT   protein deduced from cloned cDNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:7683-7687(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=93162643; PubMed=8432527;
RA   Vetter U., Vogel W., Just W., Young M.F., Fisher L.W.;
RT   "Human decorin gene: intron-exon junctions and chromosomal
RT   localization.";
RL   Genomics 15:161-168(1993).
RN   [3]
RP   SEQUENCE OF 1-70 FROM N.A.
RX   MEDLINE=93162642; PubMed=8432526;
RA   Danielson K.G., Fazzio A., Cohen I.R., Cannizzaro L., Iozzo R.V.;
RT   "The human decorin gene: intron-exon organization, discovery of two
RT   alternatively spliced exons in the 5' untranslated region, and
RT   mapping of the gene to chromosome 12q23.";
RL   Genomics 15:146-160(1993).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORMS A; B; C; D AND E).
RA   Cs-Szabo G., Glant T.T.;
RT   "Alternative splicing of human decorin.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 31-50.
RX   MEDLINE=90073579; PubMed=2590169;
RA   Roughley P.J., White R.J.;
RT   "Dermatan sulphate proteoglycans of human articular cartilage. The
RT   properties of dermatan sulphate proteoglycans I and II.";
RL   Biochem. J. 262:823-827(1989).
RN   [6]
RP   SEQUENCE OF 31-49.
RX   MEDLINE=87250639; PubMed=3597437;
RA   Fisher L.W., Hawkins G.R., Tuross N., Termine J.D.;
RT   "Purification and partial characterization of small proteoglycans I
RT   and II, bone sialoproteins I and II, and osteonectin from the mineral
RT   compartment of developing human bone.";
RL   J. Biol. Chem. 262:9702-9708(1987).
CC   -!- FUNCTION: BINDS TO TYPE I AND TYPE II COLLAGEN AND AFFECTS THE
CC       RATE OF FIBRILS FORMATION. ALSO BINDS TO FIBRONECTIN AND TGF-
CC       BETA.
CC   -!- PTM: THE GLYCOSAMINOGLYCAN CHAIN ATTACHED TO DECORIN CAN BE EITHER
CC       CHONDROITIN SULFATE OR DERMATAN SULFATE DEPENDING UPON THE
CC       TISSUE OF ORIGIN.
CC   -!- ALTERNATIVE PRODUCTS: 5 isoforms; A (shown here), B, C, D and E;
CC       are produced by alternative splicing.
```

```
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 10 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M14219; AAB00774.1; -.
DR   EMBL; L01131; AAA52301.1; ALT_SEQ.
DR   EMBL; L01125; AAA52301.1; JOINED.
DR   EMBL; L01126; AAA52301.1; JOINED.
DR   EMBL; L01127; AAA52301.1; JOINED.
DR   EMBL; L01129; AAA52301.1; JOINED.
DR   EMBL; L01130; AAA52301.1; JOINED.
DR   EMBL; M98262; AAB60901.1; -.
DR   EMBL; AF138300; AAD44713.1; -.
DR   EMBL; AF138301; AAF61437.1; -.
DR   EMBL; AF138302; AAD44714.1; -.
DR   EMBL; AF138303; AAF61438.1; -.
DR   EMBL; AF138304; AAD44715.1; -.
DR   PIR; A26476; NBHUC8.
DR   PIR; S05640; S05640.
DR   PIR; B28457; B28457.
DR   PIR; A45016; A45016.
DR   MIM; 125255; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 9.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00370; LRR; 3.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 1.
KW   Glycoprotein; Connective tissue; Extracellular matrix; Proteoglycan;
KW   Repeat; Leucine-rich repeat; Signal; Alternative splicing;
KW   Polymorphism.
FT   SIGNAL       1     16       POTENTIAL.
FT   PROPEP      17     30
FT   CHAIN       31    359       BONE PROTEOGLYCAN II.
FT   REPEAT      77     98       LRR 1.
FT   REPEAT      99    122       LRR 2.
FT   REPEAT     123    145       LRR 3.
FT   REPEAT     146    167       LRR 4.
FT   REPEAT     168    193       LRR 5.
FT   REPEAT     194    217       LRR 6.
FT   REPEAT     218    238       LRR 7.
FT   REPEAT     239    262       LRR 8.
FT   REPEAT     263    285       LRR 9.
FT   REPEAT     286    308       LRR 10.
FT   CARBOHYD    34     34       O-LINKED (GLYCOSAMINOGLYCAN).
FT   CARBOHYD   211    211       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   262    262       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   303    303       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    54     67       BY SIMILARITY.
FT   DISULFID   313    346       BY SIMILARITY.
FT   VARSPLIC    71    179       MISSING (IN ISOFORM B).
FT   VARSPLIC    73    219       MISSING (IN ISOFORM C).
FT   VARSPLIC   109    295       MISSING (IN ISOFORM D).
FT   VARSPLIC    72     75       LDKV -> CLPS (IN ISOFORM E).
FT   VARSPLIC    76    359       MISSING (IN ISOFORM E).
FT   VARIANT    273    273       E -> Q (IN DBSNP:1803344).
FT                               /FTId=VAR_011975.
FT   CONFLICT    37     37       G -> A (IN REF. 6).
FT   CONFLICT    45     45       D -> P (IN REF. 6).
SQ   SEQUENCE   359 AA;  39746 MW;  FF511E871A1A52DD CRC64;
```

```
  Query Match           12.0%;  Score 302;  DB 1;  Length 359;
  Best Local Similarity  33.1%;  Pred. No. 1.6e-14;
  Matches   85;  Conservative   41;  Mismatches  121;  Indels   10;  Gaps    6;

Qy       3 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQ 62
           ||  | |      | |    | :||  :  :| ||:|  |:|  ||   ||:|::|  |
Db      54 CPFRCQC--HLRVVQCSDLGLDKVPKDLPPDTTLLDLQNNKITEIKDGDFKNLKNLHALI 111

Qy      63 LSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSY 122
           |  | |  :  |||  |  |  | |  |:|  :|        |:||    | |  :
Db     112 LVNNKISKVSPGAFTPLVKLERLYLSKNQLKELPE---KMPKTLQELRAHENEITKVRKV 168

Qy     123 AFNRIPSLRRLDLG--ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
           || :  :  ::||   ||  | |  |||:|:  | |: :|  |:  ||   |   | ||
Db     169 TFNGLNQMIVIELGTNPLKS-SGIENGAFQGMKKLSYIRIADTNITSIPQGLP-PSLTEL 226

Qy     181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
           | || :| :  | :|| :| || :  : |  ::  :  |   | |::| :| || :|
Db     227 HLDGNKISRVDAASLKGLNNLAKLGLSFNSISAVDNGSLANTPHLRELHLDNNKLTRVPG 286

Qy     241 DLFTPLHHLERIHLHHN 257
           |     ::: ::||:|
Db     287 GL-AEHKYIQVVYLHNN 302
```

```
RESULT  12
PGS2_PIG
ID   PGS2_PIG       STANDARD;      PRT;   360 AA.
AC   Q9XSD9; Q9XSH4;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Bone proteoglycan II precursor (PG-S2) (Decorin).
GN   DCN.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A. (LONG FORM).
RC   STRAIN=YORKSHIRE;
RA   Stephenson S., Schnoke M., Vesely I.;
RT   "Cloning of the porcine decorin gene.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A. (SHORT FORM).
RC   STRAIN=YORKSHIRE; TISSUE=Aorta;
RA   Stephenson S., Schnoke M., Vesely I.;
RT   "Alternatively spliced version of the porcine decorin gene.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS TO TYPE I AND TYPE II COLLAGEN AND AFFECTS THE
CC       RATE OF FIBRILS FORMATION. ALSO BINDS TO FIBRONECTIN AND TGF-
CC       BETA (BY SIMILARITY).
CC   -!- PTM: THE GLYCOSAMINOGLYCAN CHAIN ATTACHED TO DECORIN CAN BE EITHER
CC       CHONDROITIN SULFATE OR DERMATAN SULFATE DEPENDING UPON THE
CC       TISSUE OF ORIGIN.
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 10 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF125537; AAD23578.1; -.
DR   EMBL; AF140270; AAD33862.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
```

```
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 9.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00370; LRR; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 2.
KW   Glycoprotein; Connective tissue; Extracellular matrix; Proteoglycan;
KW   Repeat; Leucine-rich repeat; Signal; Alternative splicing.
FT   SIGNAL        1     16       POTENTIAL.
FT   PROPEP       17     30       BY SIMILARITY.
FT   CHAIN        31    360       BONE PROTEOGLYCAN II.
FT   REPEAT       78     99       LRR 1.
FT   REPEAT      100    123       LRR 2.
FT   REPEAT      124    146       LRR 3.
FT   REPEAT      147    168       LRR 4.
FT   REPEAT      169    194       LRR 5.
FT   REPEAT      195    218       LRR 6.
FT   REPEAT      219    239       LRR 7.
FT   REPEAT      240    263       LRR 8.
FT   REPEAT      264    286       LRR 9.
FT   REPEAT      287    309       LRR 10.
FT   CARBOHYD     34     34       O-LINKED (GLYCOSAMINOGLYCAN) (BY
FT                                SIMILARITY).
FT   CARBOHYD    212    212       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    263    263       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    304    304       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID     55     68       BY SIMILARITY.
FT   DISULFID    314    347       POTENTIAL.
FT   VARSPLIC    281    318       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   360 AA;  39899 MW;  8573DE8DDEBA7509 CRC64;
```

```
  Query Match             11.7%;  Score 295;  DB 1;  Length 360;
  Best Local Similarity   30.5%;  Pred. No. 5.2e-14;
  Matches   97;  Conservative   44;  Mismatches  139;  Indels   38;  Gaps   10;

Qy       3 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQ 62
           ||  | |       | |    | :||  :  :| ||:|  |:|  ||   ||:|::|  |
Db      55 CPFRCQC--HLRVVQCSDLGLDKVPKDLPPDTALLDLQNNKITEIKDGDFKNLKNLHTLI 112

Qy      63 LSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSY 122
           |  | |  |  |||  |  |  | |  |:|  :|         |:|| :  | |  :
Db     113 LINNKISKISPGAFAPLVKLERLYLSKNQLKELPE---KMPKTLQELRVHENEITKVRKA 169

Qy     123 AFNRIPSLRRLDLG--ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
           || :  :  ::||   ||  | |  |||:|:  | |: :|  |:  ||   |   | ||
Db     170 VFNGLNQMIVVELGTNPLKS-SGIENGAFQGMKKLSYIRIADTNITTIPQGLP-PSLTEL 227

Qy     181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
           | || :| :   | :|| :| || :  : |  ::  :  |   | |::| :| |  :|
Db     228 HLDGNKISKVDAASLKGLNNLAKLGLGFNSISTVDNGSLANTPHLRELHLNNNKLNKVPG 287

Qy     241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIG-- 298
           |     ::: ::||:|  | :              ||  |    ||    |  | |
Db     288 GL-AEHKYIQVVYLHNN--NISA-----------VGSNDFCPPGYNTK---KASYSGVS 329

Qy     299 ---------ELDQNYFTC 307
                   |:  : | |
Db     330 LFSNPVQYWEIQPSTFRC 347
```

```
RESULT  13
PGS2_HORSE
ID   PGS2_HORSE     STANDARD;      PRT;   360 AA.
AC   O46542;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Bone proteoglycan II precursor (PG-S2) (Decorin) (Dermatan sulfate
DE   proteoglycan II) (DS-PGII).
GN   DCN.
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Richardson D.W., Dodge G.R.;
RT   "Effects of interleukin-1 beta and tumor necrosis factor-alpha on the
RT   expression of matrix related genes in cultured equine articular
RT   chondrocytes.";
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS TO TYPE I AND TYPE II COLLAGEN AND AFFECTS THE
CC       RATE OF FIBRILS FORMATION. ALSO BINDS TO FIBRONECTIN AND TGF-
CC       BETA (BY SIMILARITY).
CC   -!- PTM: THE GLYCOSAMINOGLYCAN CHAIN ATTACHED TO DECORIN CAN BE EITHER
CC       CHONDROITIN SULFATE OR DERMATAN SULFATE DEPENDING UPON THE
CC       TISSUE OF ORIGIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 10 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL-outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF038127; AAB92652.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   Pfam; PF00560; LRR; 9.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00370; LRR; 3.
DR   SMART; SM00013; LRRNT; 1.
KW   Glycoprotein; Connective tissue; Extracellular matrix; Proteoglycan;
KW   Repeat; Leucine-rich repeat; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   PROPEP       17     30       BY SIMILARITY.
FT   CHAIN        31    360       BONE PROTEOGLYCAN II.
FT   REPEAT       78     99       LRR 1.
FT   REPEAT      100    123       LRR 2.
FT   REPEAT      124    146       LRR 3.
FT   REPEAT      147    168       LRR 4.
FT   REPEAT      169    194       LRR 5.
FT   REPEAT      195    218       LRR 6.
FT   REPEAT      219    239       LRR 7.
FT   REPEAT      240    263       LRR 8.
FT   REPEAT      264    286       LRR 9.
FT   REPEAT      287    309       LRR 10.
FT   CARBOHYD     34     34       O-LINKED (GLYCOSAMINOGLYCAN) (BY
FT                                SIMILARITY).
FT   CARBOHYD    190    190       O-LINKED (GLYCOSAMINOGLYCAN) (BY
FT                                SIMILARITY).
FT   CARBOHYD    212    212       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    263    263       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    304    304       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID     55     68       BY SIMILARITY.
FT   DISULFID    314    347       BY SIMILARITY.
SQ   SEQUENCE   360 AA;  39939 MW;  2DAE97CDE16F7C45 CRC64;
```

```
  Query Match             11.6%;  Score 294;  DB 1;  Length 360;
  Best Local Similarity   30.4%;  Pred. No. 6.1e-14;
  Matches   97;  Conservative   44;  Mismatches  138;  Indels   40;  Gaps   11;

Qy       3 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQ 62
           ||  | |       | |    | :||  :  :| ||:|  |:|  ||   ||:|::|  |
Db      55 CPFRCQC--HLRVVQCSDLGLDKVPKDLPPDTTLLDLQNNKITEIKDGDFKNLKNLHALI 112

Qy      63 LSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSY 122
           |  | |  |  |||  |  |  | |  | |  :|         |:|| :  | |  :
Db     113 LVNNKISKISPGAFTPLVKLERLYLSKNHLKELPE---KMPKTLQELRVHENEITKVRKA 169
```

```
Qy    123 AFNRIPSLRRLDLG--ELKRLSYISEGAFEGLSNLRYLNLAMCNLREI-PNLTPLIKLDE 179
          || :  :  ::||   ||  | |  |||:|:  | |: :|  |:  | | | |   | |
Db    170 VFNGLNQMIVVELGTNPLKS-SGIENGAFQGMKKLSYIRIADTNITTIPPGLPP--SLTE 226

Qy    180 LDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLP 239
          | | || :: :   | :|| :| || :  : |  ::  :  |   | |::| :| |  :|
Db    227 LHLDGNKITKVDAASLRGLNNLAKLGLSFNSISAVDNGSLANTPHLRELHLDNNKLIKVP 286

Qy    240 HDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIG- 298
            |     ::: ::||:|  | :              ||  |    ||    |  | |
Db    287 GGL-ADHKYIQVVYLHNN--NISA-----------VGSNDFCPPGYNTK---KASYSGV 328

Qy    299 ----------ELDQNYFTC 307
                   |:  : | |
Db    329 SLFSNPVQYWEIQPSTFRC 347


RESULT  14
PGS2_BOVIN
ID   PGS2_BOVIN     STANDARD;      PRT;   360 AA.
AC   P21793;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Bone proteoglycan II precursor (PG-S2) (Decorin).
GN   DCN.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88133946; PubMed=3435485;
RA   Day A.A., McQuillan C.I., Termine J.D., Young M.R.;
RT   "Molecular cloning and sequence analysis of the cDNA for small
RT   proteoglycan II of bovine bone.";
RL   Biochem. J. 248:801-805(1987).
RN   [2]
RP   SEQUENCE OF 31-54.
RX   MEDLINE=89123388; PubMed=2914936;
RA   Choi H.U., Johnson T.L., Pal S., Tang L.H., Rosenberg L., Neame P.J.;
RT   "Characterization of the dermatan sulfate proteoglycans, DS-PGI and
RT   DS-PGII, from bovine articular cartilage and skin isolated by octyl-
RT   sepharose chromatography.";
RL   J. Biol. Chem. 264:2876-2884(1989).
CC   -!- FUNCTION: BINDS TO TYPE I AND TYPE II COLLAGEN AND AFFECTS THE
CC       RATE OF FIBRILS FORMATION. ALSO BINDS TO FIBRONECTIN AND TGF-
CC       BETA.
CC   -!- PTM: THE GLYCOSAMINOGLYCAN CHAIN ATTACHED TO DECORIN CAN BE EITHER
CC       CHONDROITIN SULFATE OR DERMATAN SULFATE DEPENDING UPON THE
CC       TISSUE OF ORIGIN.
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 10 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y00712; CAA68702.1; -.
DR   PIR; S06280; S06280.
DR   PIR; B31430; B31430.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 9.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00370; LRR; 2.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 2.
KW   Glycoprotein; Connective tissue; Extracellular matrix; Proteoglycan;
KW   Repeat; Leucine-rich repeat; Signal.
FT   SIGNAL       1    16       POTENTIAL.
FT   PROPEP      17    30
FT   CHAIN       31   360       BONE PROTEOGLYCAN II.
FT   REPEAT      78    99       LRR 1.
FT   REPEAT     100   123       LRR 2.
FT   REPEAT     124   146       LRR 3.
FT   REPEAT     147   168       LRR 4.
FT   REPEAT     169   194       LRR 5.
FT   REPEAT     195   218       LRR 6.
FT   REPEAT     219   239       LRR 7.
FT   REPEAT     240   263       LRR 8.
FT   REPEAT     264   286       LRR 9.
FT   REPEAT     287   309       LRR 10.
FT   CARBOHYD    34    34       O-LINKED (GLYCOSAMINOGLYCAN) (BY
FT                              SIMILARITY).
FT   CARBOHYD   212   212       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   263   263       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   304   304       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID    55    68       BY SIMILARITY.
FT   DISULFID   314   347       BY SIMILARITY.
SQ   SEQUENCE   360 AA;  39837 MW;  71E84DA2D87552C0 CRC64;


  Query Match             11.6%;  Score 293;  DB 1;  Length 360;
  Best Local Similarity   29.9%;  Pred. No. 7.2e-14;
  Matches   95;  Conservative   46;  Mismatches  139;  Indels   38;  Gaps   10;

Qy      3 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQ 62
            ||  | |       | |    | :||  :  :| ||:|  |:|  ||   ||:|::|  |
Db     55 CPFRCQC--HLRVVQCSDLGLEKVPKDLPPDTALLDLQNNKITEIKDGDFKNLKNLHTLI 112

Qy     63 LSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSY 122
          |  | |  |  |||  |  |  | |  |:|  :|         |:|| :  | |  :
Db    113 LINNKISKISPGAFAPLVKLERLYLSKNQLKELPE---KMPKTLQELRVHENEITKVRKS 169

Qy    123 AFNRIPSLRRLDLG--ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
          || :  :  ::||   ||  | |  |||:|:  | |: :|  |:  ||   |  | ||
Db    170 VFNGLNQMIVVELGTNPLKS-SGIENGAFQGMKKLSYIRIADTNITTIPQGLP-PSLTEL 227

Qy    181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
          | || :: :   | :|| :| || :  : |  ::  :  |   | |::| :| |  :|
Db    228 HLDGNKITKVDAASLKGLNNLAKLGLSFNSISAVDNGSLANTPHLRELHLNNNKLAKVPG 287

Qy    241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIG-- 298
          :     ::: ::||:|    |   :            ||  |    ||    |  | |
Db    288 GV-ADHKYIQVVYLHNN----NISAI----------GSNDFCPPGYNTK---KASYSGVS 329

Qy    299 ---------ELDQNYFTC 307
                  |:  : | |
Db    330 LFSNPVQYWEIQPSTFRC 347


RESULT  15
PGS2_CHICK
ID   PGS2_CHICK     STANDARD;      PRT;   357 AA.
AC   P28675;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Bone proteoglycan II precursor (PG-S2) (Decorin).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
```

```
RC   STRAIN=WHITE LEGHORN; TISSUE=Cornea;
RX   MEDLINE=92296755; PubMed=1605630;
RA   Li W., Vergnes J.P., Cornuet P.K., Hassel J.R.;
RT   "cDNA clone to chick corneal chondroitin/dermatan sulfate
RT   proteoglycan reveals identity to decorin.";
RL   Arch. Biochem. Biophys. 296:190-197(1992).
CC   -!- FUNCTION: BINDS TO TYPE I AND TYPE II COLLAGEN AND AFFECTS THE
CC       RATE OF FIBRILS FORMATION. ALSO BINDS TO FIBRONECTIN AND TGF-
CC       BETA.
CC   -!- PTM: THE GLYCOSAMINOGLYCAN CHAIN ATTACHED TO DECORIN CAN BE EITHER
CC       CHONDROITIN SULFATE OR DERMATAN SULFATE DEPENDING UPON THE
CC       TISSUE OF ORIGIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 10 LEUCINE-RICH REPEATS (LRR).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X63797; CAA45318.1; -.
DR   PIR; S22197; S22197.
DR   PIR; S24317; S24317.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00560; LRR; 10.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00370; LRR; 2.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 2.
KW   Glycoprotein; Connective tissue; Extracellular matrix; Proteoglycan;
KW   Repeat; Leucine-rich repeat; Signal.
FT   SIGNAL        1     16
FT   PROPEP       17     30
FT   CHAIN        31    357       BONE PROTEOGLYCAN II.
FT   REPEAT       75     96       LRR 1.
FT   REPEAT       97    120       LRR 2.
FT   REPEAT      121    143       LRR 3.
FT   REPEAT      144    165       LRR 4.
FT   REPEAT      166    191       LRR 5.
FT   REPEAT      192    215       LRR 6.
FT   REPEAT      216    236       LRR 7.
FT   REPEAT      237    260       LRR 8.
FT   REPEAT      261    283       LRR 9.
FT   REPEAT      284    306       LRR 10.
FT   CARBOHYD     34     34       O-LINKED (GLYCOSAMINOGLYCAN) (BY
FT                                SIMILARITY).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    260    260       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID     52     65       BY SIMILARITY.
FT   DISULFID    311    344       BY SIMILARITY.
SQ   SEQUENCE   357 AA;  39687 MW;  31B104C7C3CD711D CRC64;
```

```
  Query Match            11.5%;  Score 291;  DB 1;  Length 357;
  Best Local Similarity  33.1%;  Pred. No. 9.9e-14;
  Matches   85;  Conservative   37;  Mismatches  125;  Indels   10;  Gaps    6;

Qy       3 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQ 62
           ||  | |      | |    |  ||  :  :| ||:|  |:|  ||   ||:|::|  |
Db      52 CPFRCQC--HLRVVQCSDLGLERVPKDLPPDTTLLDLQNNKITEIKEGDFKNLKNLHALI 109

Qy      63 LSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSY 122
           |  | |  |   ||  |  |  | |  | |  :|        |:|:    | |  :
Db     110 LVNNKISKISPAAFAPLKKLERLYLSKNNLKELPEN---MPKSLQEIRAHENEISKLRKA 166

Qy     123 AFNRIPSLRRLDLG--ELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
           || :  :  |:||   ||  | |  |||:|:  | |: :|  |:  ||   |   | ||
Db     167 VFNGLNQVIVLELGTNPLKS-SGIENGAFQGMKRLSYIRIADTNITSIPKGLP-PSLTEL 224

Qy     181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
           | || :| |      || :| || :  : |  :|  : :|:  | |::| :| |  :|
Db     225 HLDGNKISKIDAEGLSGLTNLAKLGLSFNSISSVENGSLNNVPHLRELHLNNNELVRVPS 284

Qy     241 DLFTPLHHLERIHLHHN 257
           |     ::: ::||:|
Db     285 GL-GEHKYIQVVYLHNN 300
```

Search completed: August 26, 2002, 15:50:43
Job time: 240 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 2002, 15:46:18 ; Search time 31.39 Seconds
(without alignments)
2634.329 Million cell updates/sec

Title: US-09-905-056-292_COPY_45_522
Perfect score: 2524
Sequence: 1 QTCPSVCSCSNQFSKVICVR.....KTFTIPVTDINSGIPGIDEV 478

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2524	100.0	640	4	Q9HCJ2	Q9hcj2 homo sapien
2	1581.5	62.7	653	4	Q9HBW1	Q9hbw1 homo sapien
3	1556.5	61.7	649	4	Q96A85	Q96a85 homo sapien
4	1151.5	45.6	422	4	Q9NT99	Q9nt99 homo sapien
5	872.5	34.6	441	11	Q99PH1	Q99ph1 mus musculu
6	565.5	22.4	606	4	Q9BZ20	Q9bz20 homo sapien
7	565.5	22.4	614	4	Q96FE5	Q96fe5 homo sapien
8	560.5	22.2	614	6	Q9N008	Q9n008 macaca fasc
9	559.5	22.2	614	11	Q9D1T0	Q9d1t0 mus musculu
10	453	17.9	719	4	Q96NI6	Q96ni6 homo sapien
11	443.5	17.6	1021	5	Q9V430	Q9v430 drosophila
12	421	16.7	705	4	O43377	O43377 homo sapien
13	421	16.7	708	4	Q9H3W5	Q9h3w5 homo sapien
14	413.5	16.4	1093	4	Q96JA1	Q96ja1 homo sapien
15	413.5	16.4	1094	4	Q9BYB8	Q9byb8 homo sapien
16	410.5	16.3	730	4	Q9P231	Q9p231 homo sapien
17	410	16.2	707	11	P97860	P97860 mus musculu
18	408	16.2	789	6	Q9BE71	Q9be71 macaca fasc
19	406	16.1	832	4	Q9ULH4	Q9ulh4 homo sapien
20	403	16.0	788	11	Q9CYK3	Q9cyk3 mus musculu
21	402.5	15.9	1091	11	P70193	P70193 mus musculu
22	401.5	15.9	359	4	Q96RJ5	Q96rj5 homo sapien
23	401.5	15.9	492	11	Q99KT6	Q99kt6 mus musculu
24	398.5	15.8	716	11	Q61809	Q61809 mus musculu
25	398	15.8	718	13	O73675	O73675 xenopus lae
26	397.5	15.7	359	4	O15335	O15335 homo sapien
27	396	15.7	707	11	Q9ESY6	Q9esy6 rattus norv
28	393.5	15.6	1531	11	Q9WVB5	Q9wvb5 mus musculu
29	389.5	15.4	1515	13	Q9DE37	Q9de37 brachydanio
30	389.5	15.4	1531	11	O88279	O88279 rattus norv
31	389	15.4	1521	11	Q9R1B9	Q9r1b9 mus musculu
32	388.5	15.4	358	11	O55226	O55226 mus musculu
33	388.5	15.4	674	4	Q9NZU1	Q9nzu1 homo sapien
34	388	15.4	1521	4	O95710	O95710 homo sapien
35	388	15.4	1529	4	O94813	O94813 homo sapien
36	387.5	15.4	1530	11	Q9WUG5	Q9wug5 rattus norv
37	386	15.3	1525	4	Q9Y5Q7	Q9y5q7 homo sapien
38	384.5	15.2	1496	4	Q92626	Q92626 homo sapien
39	382.5	15.2	358	11	O70210	O70210 rattus norv
40	380	15.1	628	4	Q9BTN0	Q9btn0 homo sapien
41	378	15.0	796	11	Q9WVC1	Q9wvc1 rattus norv
42	377.5	15.0	1173	5	Q9V7J8	Q9v7j8 drosophila
43	377	14.9	1534	4	O75093	O75093 homo sapien
44	376	14.9	1065	4	O94898	O94898 homo sapien
45	375	14.9	660	4	O43155	O43155 homo sapien

ALIGNMENTS

RESULT 1
Q9HCJ2
ID Q9HCJ2 PRELIMINARY; PRT; 640 AA.
AC Q9HCJ2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA1580 PROTEIN (FRAGMENT).
GN KIAA1580.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20450683; PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirosawa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVIII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro.";
RL DNA Res. 7:273-281(2000).
DR EMBL; AB046800; BAB13406.1; -.
DR HSSP; P22888; 1LUT.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003592; LRR_out.
DR InterPro; IPR003591; LRR_typ.
DR Pfam; PF00047; ig; 1.
DR Pfam; PF00560; LRR; 9.
DR Pfam; PF01463; LRRCT; 1.
DR Pfam; PF01462; LRRNT; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00409; IG; 1.
DR SMART; SM00408; IGc2; 1.

```
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00370; LRR; 6.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 8.
KW   Immunoglobulin domain.
FT   NON_TER       1       1
SQ   SEQUENCE   640 AA;  71949 MW;  6058974872636838 CRC64;


  Query Match             100.0%;  Score 2524;  DB 4;  Length 640;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-181;
  Matches  478;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     45 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 104

Qy     61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    105 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 164

Qy    121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    165 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 224

Qy    181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    225 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 284

Qy    241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGEL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    285 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGEL 344

Qy    301 DQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    345 DQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIA 404

Qy    361 VLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    405 VLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEP 464

Qy    421 SQDEARTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEV 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    465 SQDEARTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTDINSGIPGIDEV 522


RESULT   2
Q9HBW1
ID   Q9HBW1      PRELIMINARY;      PRT;   653 AA.
AC   Q9HBW1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BRAIN TUMOR ASSOCIATED PROTEIN NAG14.
GN   NAG14.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wang J., Bin L., Jiang N., Li G.;
RT   "Homo sapiens brain-specific gene (BAG), downregulated in brain tumor,
RT   mRNA.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF196976; AAG28019.2; -.
DR   HSSP; P23945; 1XUN.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 8.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00370; LRR; 3.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 7.
KW   Immunoglobulin domain.
SQ   SEQUENCE   653 AA;  72717 MW;  38159C81F6850E37 CRC64;


  Query Match             62.7%;  Score 1581.5;  DB 4;  Length 653;
  Best Local Similarity   60.9%;  Pred. No. 1.3e-110;
  Matches  298;  Conservative   77;  Mismatches   93;  Indels   21;  Gaps    8;

Qy      1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
          | ||||||||||||||:| |: | ||| || :||| ||| || ||:|: ::|:|| |||:
Db     44 QNCPSVCSCSNQFSKVVCTRRGLSEVPQGIPSNTRYLNLMENNIQMIQADTFRHLHHLEV 103

Qy     61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
          ||| || || ||:|||||||:|||||||| || ||:||| ||||:||||||||||||
Db    104 LQLGRNSIRQIEVGAFNGLASLNTLELFDNWLTVIPSGAFEYLSKLRELWLRNNPIESIP 163

Qy    121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
          ||||||:||| ||||||||:| |||||||||| ||:|||| |||:::::||||||: |:||
Db    164 SYAFNRVPSLMRLDLGELKKLEYISEGAFEGLFNLKYLNLGMCNIKDMPNLTPLVGLEEL 223

Qy    181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
          ::|||   |||||| ||  |:|||:: ||: :||||||| | ||||:||||||: |||
Db    224 EMSGNHFPEIRPGSFHGLSSLKKLWVMNSQVSLIERNAFDGLASLVELNLAHNNLSSLPH 283

Qy    241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGEL 300
          |||||| :|  :|||||||||:||||||:||:::  |:|: || || ||: | :::|||: |:
Db    284 DLFTPLRYLVELHLHHNPWNCDCDILWLAWWLREYIPTNSTCCGRCHAPMHMRGRYLVEV 343

Qy    301 DQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIA 360
          ||  | | || |:: | |||::|| ||||||  :  ::|| |: |||||::| :   ||:
Db    344 DQASFQCSAPFIMDAPRDLNISEGRMAELKCR-TPPMSSVKWLLPNGTVLSHASRHPRIS 402

Qy    361 VLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAA--TTTPFSYFSTVTVETM 418
          ||:||||||::| : |||:|||||:|  ||: ||| |||: |    |: :|:|:||||||
Db    403 VLNDGTLNFSHVLLSDTGVYTCMVTNVAGNSNASAYLNVSTAELNTSNYSFFTTVTVETT 462

Qy    419 EPSQDEARTTDNNVGPTPVVDWETTNV------TTSLTP--QSTRSTEKTFTIPVTDINS 470
          | | ::   |     | |     ||:       ||| |   |:||   |   :| ||
Db    463 EISPED---TTRKYKPVP-----TTSTGYQPAYTTSTTVLIQTTR-VPKQVAVPATDTTD 513

Qy    471 GI-PGIDEV 478
            :   :|||
Db    514 KMQTSLDEV 522


RESULT   3
Q96A85
ID   Q96A85      PRELIMINARY;      PRT;   649 AA.
AC   Q96A85;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 72.3 KDA PROTEIN.
GN   BAG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wang J.;
RL   Thesis (2000), Zhongshan Medical University, Guangzhou, China.
DR   EMBL; AJ297858; CAC82651.1; -.
KW   Signal; Hypothetical protein.
FT   SIGNAL       1     38       POTENTIAL.
SQ   SEQUENCE   649 AA;  72313 MW;  9C11C5ABC7E536CF CRC64;


  Query Match             61.7%;  Score 1556.5;  DB 4;  Length 649;
  Best Local Similarity   60.5%;  Pred. No. 9.4e-109;
  Matches  296;  Conservative   75;  Mismatches   93;  Indels   25;  Gaps    9;

Qy       1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
           | |||||||||||||:| |: | ||| || :||| ||| || ||:|: ::|:|| |||:
Db      44 QNCPSVCSCSNQFSKVVCTRRGLSEVPQGIPSNTRYLNLMENNIQMIQADTFRHLHHLEV 103

Qy      61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
           ||| || || ||:|||||||:||||||||| || ||:||| |||||:||||||||||||||
Db     104 LQLGRNSIRQIEVGAFNGLASLNTLELFDNWLTVIPSGAFEYLSKLRELWLRNNPIESIP 163

Qy     121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDEL 180
           ||||||:||| |||||||:| |||||||||| ||:|||| |||:::::|||||: |:||
Db     164 SYAFNRVPSLMRLDLGELKKLEYISEGAFEGLFNLKYLNLGMCNIKDMPNLTPLVGLEEL 223

Qy     181 DLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPH 240
           ::|||   |||||| ||  |:|||:: |     |||||| | ||||:|||||||: |||
Db     224 EMSGNHFPEIRPGSFHGLSSLKKLWVMNSH----ERNAFDGLASLVELNLAHNNLSSLPH 279

Qy     241 DLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGEL 300
           |||||| :|  :|||||||||:||||||:||:::  |:|: || ||: | :::|||: |:
Db     280 DLFTPLRYLVELHLHHNPWNCDCDILWLAWWLREYIPTNSTCCGRCHAPMHMRGRYLVEV 339

Qy     301 DQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIA 360
           ||  | | || |:: | |||::||  |||||| :  ::|| |: |||||::| :   ||:
Db     340 DQASFQCSAPFIMDAPRDLNISEGRMAELKCR-TPPMSSVKWLLPNGTVLSHASRHPRIS 398

Qy     361 VLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAA--TTTPFSYFSTVTVETM 418
           ||:||||||::| : |||:|||||:|  ||: ||| |||: |   |: :|:|:|:||||||
Db     399 VLNDGTLNFSHVLLSDTGVYTCMVTNVAGNSNASAYLNVSTAELNTSNYSFFTTVTVETT 458

Qy     419 EPSQDEARTTDNNVGPTPVVDWETTNV------TTSLTP--QSTRSTEKTFTIPVTDINS 470
           | | ::   |    | |    ||:      ||| |   |:||   |   :| ||
Db     459 EISPED---TTRKYKPVP-----TTSTGYQPAYTTSTTVLIQTTR-VPKQVAVPATDTTD 509

Qy     471 GI-PGIDEV 478
             :   :|||
Db     510 KMQTSLDEV 518


RESULT   4
Q9NT99
ID   Q9NT99      PRELIMINARY;      PRT;   422 AA.
AC   Q9NT99;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 45.1 KDA PROTEIN (FRAGMENT).
GN   DKFZP761A179.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=AMYGDALA;
RA   Blum H., Bauersachs S., Mewes H.W., Gassenhuber J., Wiemann S.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL137451; CAB70743.1; -.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   InterPro; IPR000217; Tubulin.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 3.
DR   Pfam; PF01463; LRRCT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00370; LRR; 1.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00369; LRR_TYP; 2.
DR   PROSITE; PS00227; TUBULIN; UNKNOWN_1.
KW   Hypothetical protein; Immunoglobulin domain.
FT   NON_TER       1       1
FT   NON_TER     422     422
SQ   SEQUENCE   422 AA;  45110 MW;  164C6AF1A4BC440C CRC64;


  Query Match             45.6%;  Score 1151.5;  DB 4;  Length 422;
  Best Local Similarity   59.8%;  Pred. No. 1.5e-78;
  Matches  213;  Conservative   53;  Mismatches   51;  Indels   39;  Gaps    6;

Qy     162 MCNLREIPNLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDN 221
           ||||::|||| |::|:||:|||| |  |||||||||  |:|||:: :|:  |||||||:
Db       1 MCNLKDIPNLTALVRLEELELSGNRLDLIRPGSFQGLTSLRKLWLMHAQVATIERNAFDD 60

Qy     222 LQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTA 281
           |:|| |:||:|||| ||||||||| |||:||:|||||:||||:||||||:|:  ||||
Db      61 LKSLEELNLSHNNLMSLPHDLFTPLHRLERVHLNHNPWHCNCDVLWLSWWLKETVPSNTT 120

Qy     282 CCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVS 341
           |||||: |  |||||||||||::||||||||||| |||||||||||||  ||:|||:
Db     121 CCARCHAPAGLKGRYIGELDQSHFTCYAPVIVEPPTDLNVTEGMAAELKCRTGTSMTSVN 180

Qy     342 WITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTA 401
           |:|||||:||||:|:|||:|| |||||||||||||| |||||:|| ||||||||||:|
Db     181 WLTPNGTLMTHGSYRVRISVLHDGTLNFTNVTVQDTGQYTCMVTNSAGNTTASATLNVSA 240

Qy     402 ATTTP----------------------FSYFSTVTVETME--PSQD--EARTTDNN-VG 433
                                        ::||:|||||:|  | ::  : | |:    |
Db     241 VDPVAAGGTGSGGGGPGGSGGVGGGSGGYTYFTTVTVETLETQPGEEALQPRGTEKEPPG 300

Qy     434 PTPVVDW----------ETTNVTTSLTPQSTRSTEKTFTIPVTDI-NSGIPGIDEV 478
           ||    |            :: ||:  |:|:| ||| ||:|:||:  : :  :|:|
Db     301 PTTDGVWGGGRPGDAAGPASSSTTAPAPRSSRPTEKAFTVPITDVTENALKDLDDV 356


RESULT   5
Q99PH1
ID   Q99PH1      PRELIMINARY;      PRT;   441 AA.
AC   Q99PH1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LIBG-LIKE PROTEIN (FRAGMENT).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wang J., Bin L., Li G.;
RT   "Brain tumor-associated mRNA.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF290542; AAG60620.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR003592; LRR_out.
```

```
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 2.
DR   Pfam; PF01463; LRRCT; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00370; LRR; 2.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00369; LRR_TYP; 3.
KW   Immunoglobulin domain.
FT   NON_TER       1       1
SQ   SEQUENCE   441 AA;  48844 MW;  E2D17AB4607AC475 CRC64;


  Query Match             34.6%;  Score 872.5;  DB 11;  Length 441;
  Best Local Similarity   52.2%;  Pred. No. 1.4e-57;
  Matches  163;  Conservative   57;  Mismatches   87;  Indels    5;  Gaps    4;

Qy    170 NLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEIN 229
          |||||: |:||::|||   ||||| ||  |:|||:: ||: :|||||| | ||||:|
Db      1 NLTPLVGLEELEMSGNHFPEIRPGSFHGLSSLKKLWVMNSQVSLIERNAFDGLASLVELN 60

Qy    230 LAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTP 289
          ||||||: ||||||||| :|  :|||||||||:||||||:||:::  |:|: || ||: |
Db     61 LAHNNLSSLPHDLFTPLRYLVELHLHHNPWNCDCDILWLAWWLREYIPTNSTCCGRCHAP 120

Qy    290 PNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTV 349
          :::|||: |:||  | | || |:: | |||::|   |||||| :  ::|| |: ||||| 
Db    121 MHMRGRYLVEVDQAAFQCSAPFIMDAPRDLNISEDRMAELKCR-TPPMSSVKWLLPNGTV 179

Qy    350 MTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAA--TTTPF 407
          ::| :   ||:||:|||||: | : |||:|||||:|  ||: ||| |||::|   |  | 
Db    180 LSHASRHPRISVLNDGTLNFSRVLLIDTGVYTCMVTNVAGNSNASAYLNVSSAELNTPNF 239

Qy    408 SYFSTVTVETMEPSQDEARTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTIPVTD 467
          |:|:|||||| | | ::  |      ||    ::    |::    |    |   :| ||
Db    240 SFFTTVTVETTEISPEDI-TRKYKPVPTTSTGYQPAYTTSTTVLIQTTRVPKQVPVPSTD 298

Qy    468 INSGI-PGIDEV 478
             :   :|||
Db    299 TTDKMQTSLDEV 310


RESULT   6
Q9BZ20
ID   Q9BZ20      PRELIMINARY;      PRT;   606 AA.
AC   Q9BZ20;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BA438B23.1 (NEURONAL LEUCINE-RICH REPEAT PROTEIN) (CDNA FLJ31810 FIS,
DE   CLONE NT2RI2009289, WEAKLY SIMILAR TO CARBOXYPEPTIDASE N 83 KDA
DE   CHAIN).
GN   BA438B23.1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Babbage A.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA   Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
RA   Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA   Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA   Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA   Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL353746; CAC22713.1; -.
DR   EMBL; AK056372; BAB71167.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 11.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 10.
KW   Immunoglobulin domain.
SQ   SEQUENCE   606 AA;  68065 MW;  CB608E281B066B9D CRC64;


  Query Match             22.4%;  Score 565.5;  DB 4;  Length 606;
  Best Local Similarity   29.3%;  Pred. No. 2.6e-34;
  Matches  157;  Conservative   79;  Mismatches  196;  Indels  103;  Gaps   11;

Qy      3 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQ 62
          ||: | || |   | | |: |  :|:||   |::|:| :|::: :    |     || :
Db     28 CPARCECSAQNKSVSCHRRRLIAIPEGIPIETKILDLSKNRLKSVNPEEFISYPLLEEID 87

Qy     63 LSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSY 122
          || | |  :| |||| | || :| |  |||  :| | |  || | :| :  | |  :  |
Db     88 LSDNIIANVEPGAFNNLFNLRSLRLKGNRLKLVPLGVFTGLSNLTKLDISENKIVILLDY 147

Qy    123 AFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIP------------- 169
          |  : :|: |::|:   | |||  || || :|  | |  |||  :|
Db    148 MFQDLHNLKSLEVGD-NDLVYISHRAFSGLLSLEQLTLEKCNLTAVPTEALSHLRSLISL 206

Qy    170 ----------------------------------------------NLTP------------- 173
                                                           |||
Db    207 HLKHLNINNMPVYAFKRLFHLKHLEIDYWPLLDMMPANSLYGLNLTSLSVTNTNLSTVPF 266

Qy    174 -----LIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEI 228
               |: |  |:|| | :| |  | |  |: ||:| :: :|:: || ::|  |: | :
Db    267 LAFKHLVYLTHLNLSYNPISTIEAGMFSDLIRLQELHIVGAQLRTIEPHSFQGLRFLRVL 326

Qy    229 NLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWL-----SWWIKDMAPSNTACC 283
          |:: | |  |  ::|:    || : :::|| |:| :||:     :       |
Db    327 NVSQNLLETLEENVFSSPRALEVLSINNNPLACDCRLLWILQRQPTLQFGGQQP------ 380

Qy    284 ARCNTPPNLKGRYIGELDQN----YFTCYAPVIVEPPAD-LNVTEGMAAELKCRASTSLT 338
            |  |  :: |   :         ||||  | | |     | | ||   :|:| |
Db    381 -MCAGPDTIRERSFKDFHSTALSFYFTCKKPKIREKKLQHLLVDEGQTVQLECSADGDPQ 439

Qy    339 SV-SWITPNG---TVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTAS 394
          | ||:||     |  ::|    | || ||||       ||:||| |: ||: || | :
Db    440 PVISWVTPRRRFITTKSNG----RATVLGDGTLEIRFAQDQDSGMYVCIASNAAGNDTFT 495

Qy    395 ATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTPVVDWETTNVTTSL 449
          |:| |    :  | | :   :   : :   :  |: | |  |  :| :|  |:|::
Db    496 ASLTVKGFASDRFLYANRTPMYMTDSNDTISNGTNAN---TFSLDLKTILVSTAM 547


RESULT   7
Q96FE5
ID   Q96FE5      PRELIMINARY;      PRT;   614 AA.
AC   Q96FE5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
```

```
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   UNKNOWN (PROTEIN FOR MGC:17422).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN, AND ANAPLASTIC OLIGODENDROGLIOMA WITH 1P/19Q LOSS;
RA   Strausberg R.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC011057; AAH11057.1; -.
SQ   SEQUENCE   614 AA;  69145 MW;  EFD967E3B716698D CRC64;


  Query Match           22.4%;  Score 565.5;  DB 4;  Length 614;
  Best Local Similarity  30.5%;  Pred. No. 2.7e-34;
  Matches  162;  Conservative   74;  Mismatches  202;  Indels   93;  Gaps   12;

Qy       3 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQ 62
Db      36 CPPRCECSAQDRAVLCHRKRFVAVPEGIPTETRLLDLGKNRIKTLNQDEFASFPHLEELE 95

Qy      63 LSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSY 122
Db      96 LNENIVSAVEPGAFNNLFNLRTLGLRSNRLKLIPLGVFTGLSNLTKLDISENKIVILLDY 155

Qy     123 AFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIP------------- 169
Db     156 MFQDLYNLKSLEVGD-NDLVYISHRAFSGLNSLEQLTLEKCNLTSIPTEALSHLHGLIVL 214

Qy     170 ------------------------------------------------------NLTP--- 173
Db     215 RLRHLNINAIRDYSFKRLYRLKVLEISHWPYLDTMTPNCLYGLNLTSLSITHCNLTAVPY 274

Qy     174 -----LIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEI 228
Db     275 LAVRHLVYLRFLNLSYNPISTIEGSMLHELLRLQEIQLVGGQLAVVEPYAFRGLNYLRVL 334

Qy     229 NLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWL---SWWIKDMAPSNTACCAR 285
Db     335 NVSGNQLTTLEESVFHSVGNLETLILDSNPLACDCRLLWVFRRRWRLNFNRQQPT----- 389

Qy     286 CNTPPNLKGRYIGE----LDQNYFTCYAPVIVEPPA-DLNVTEGMAAELKCRA-STSLTS 339
Db     390 CATPEFVQGKEFKDFPDVLLPNYFTCRRARIRDRKAQQVFVDEGHTVQFVCRADGDPPPA 449

Qy     340 VSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNV 399
Db     450 ILWLSPRKHLVS-AKSNGRLTVFPDGTLEVRYAQVQDNGTYLCIAANAGGNDSMPAHLHV 508

Qy     400 TA-ATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTPVVDWETTNVTTSL 449
Db     509 RSYSPDWPHQPNKTFAFISNQPGEGEANSTRATV-PFP-FDIKTLIIATTM 557


RESULT   8
Q9N008
ID   Q9N008      PRELIMINARY;      PRT;   614 AA.
AC   Q9N008;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 69.2 KDA PROTEIN.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CEREBELLUM CORTEX;
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB046639; BAB03557.1; -.
DR   HSSP; P23945; 1XUN.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 9.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00370; LRR; 5.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 1.
KW   Hypothetical protein; Immunoglobulin domain.
SQ   SEQUENCE   614 AA;  69187 MW;  BA6C8BC7C993BE9A CRC64;


  Query Match           22.2%;  Score 560.5;  DB 6;  Length 614;
  Best Local Similarity  30.3%;  Pred. No. 6.3e-34;
  Matches  161;  Conservative   74;  Mismatches  203;  Indels   93;  Gaps   12;

Qy       3 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQ 62
Db      36 CPPRCECSAQDRAVLCHRKRFVAVPEGIPTETRLLDLGKNRIKTLNQDEFASFPHLEELE 95

Qy      63 LSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSY 122
Db      96 LNENIVSAVEPGAFNNLFNLRTLGLRSNRLKLIPLGVFTGLSNLTKLDISENKIVILLDY 155

Qy     123 AFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIP------------- 169
Db     156 MFQDLYNLKSLEVGD-NDLVYISHRAFSGLNSLEQLTLEKCNLTSIPTEALSHLHGLIVL 214

Qy     170 ------------------------------------------------------NLTP--- 173
Db     215 RLRHLNINAIRDYSFKRLYRLKVLEISHWPYLDTMTPNCLYGLNLTSLSITHCNLTAVPY 274

Qy     174 -----LIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEI 228
Db     275 LAVRHLVYLRFLNLSYNPISTIEGSMLHELLRLQEIQLVGGQLAMVEPYAFRGLNYLRVL 334

Qy     229 NLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWL---SWWIKDMAPSNTACCAR 285
Db     335 NVSGNQLTTLEESVFHSVGNLETLILDSNPLACDCRLLWVFRRRWRLNFNRQQPT----- 389

Qy     286 CNTPPNLKGRYIGE----LDQNYFTCYAPVIVEPPA-DLNVTEGMAAELKCRA-STSLTS 339
Db     390 CATPEFVQGKEFKDFPDVLLPNYFTCRRARIRDRKAQQVFVDEGHTVQFVCRADGDPPPA 449

Qy     340 VSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNV 399
Db     450 ILWLSPRKHLVS-AKSNGRLTVFPDGTLEVRYAQVQDNGTYLCIAANAGGNDSMPAHLHV 508

Qy     400 TA-ATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTPVVDWETTNVTTSL 449
Db     509 RSYSPDWPHQPNKTFAFIPNQPGEGEANSTRATV-PFP-FDIKTLIIATTM 557


RESULT   9
Q9D1T0
ID   Q9D1T0      PRELIMINARY;      PRT;   614 AA.
AC   Q9D1T0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
```

```
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ADULT MALE TESTIS CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
DE   CLONE:4930471K13, FULL INSERT SEQUENCE.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TESTIS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK027262; BAB32403.1; -.
DR   HSSP; P23945; 1XUN.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 9.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 9.
KW   Immunoglobulin domain.
SQ   SEQUENCE   614 AA;  69100 MW;  41CFF40C21335681 CRC64;
```

```
  Query Match           22.2%;  Score 559.5;  DB 11;  Length 614;
  Best Local Similarity  30.3%;  Pred. No. 7.5e-34;
  Matches  161;  Conservative   73;  Mismatches  204;  Indels   93;  Gaps   12;

Qy       3 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQ 62
Db      36 CPPRCECSAQDRAVLCHRKRFVAVPEGIPTETRLLDLGKNRIKTLNQDEFASFPHLEELE 95

Qy      63 LSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSY 122
Db      96 LNENIVSAVEPGAFNNLFNLRTLGLRSNRLKLIPLGVFTGLSNLTKLDISENKIVILLDY 155

Qy     123 AFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIP------------- 169
Db     156 MFQDLYNLKSLEVGD-NDLVYISHRAFSGLNSLEQLTLEKCNLTSIPTEALSHLHGLIVL 214

Qy     170 ----------------------------------------------------NLTP--- 173
```

```
Db     215 RLRHLNINAIRDYSFKRLYRLKVLEISHWPYLDTMTPNCLYGLNLTSLSITHCNLTAVPY 274

Qy     174 -----LIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEI 228
Db     275 LAVRHLVYLRFLNLSYNPIGTIEGSMLHELLRLQEIQLVGGQLAVVEPYAFRGLNYLRVL 334

Qy     229 NLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWL---SWWIKDMAPSNTACCAR 285
Db     335 NVSGNQLTTLEESAFHSVGNLETLILDSNPLACDCRLLWVFRRRWRLNFNRQQPT----- 389

Qy     286 CNTPPNLKGRYIGE----LDQNYFTCYAPVIVEPPA-DLNVTEGMAAELKCRA-STSLTS 339
Db     390 CATPEFVQGKEFKDFPDVLLPNYFTCRRAHIRDRKAQQVFVDEGHTVQFVCRADGDPPPA 449

Qy     340 VSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNV 399
Db     450 ILWLSPRKHLVS-AKSNGRLTVFPDGTLEVRYAQVQDNGTYLCIAANAGGNDSMPAHLHV 508

Qy     400 TA-ATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTPVVDWETTNVTTSL 449
Db     509 RSYSPDWPHQPNKTFAFISNQPGEGEANSTRATV-PFP-FDIKTLIIATTM 557
```

```
RESULT  10
Q96NI6
ID   Q96NI6      PRELIMINARY;      PRT;   719 AA.
AC   Q96NI6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CDNA FLJ30803 FIS, CLONE FEBRA2001245, WEAKLY SIMILAR TO NAG14.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA   Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA   Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA   Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK055365; BAB70910.1; -.
SQ   SEQUENCE   719 AA;  79472 MW;  CE301F219705FF4B CRC64;
```

```
  Query Match           17.9%;  Score 453;  DB 4;  Length 719;
  Best Local Similarity  27.7%;  Pred. No. 9.4e-26;
  Matches  137;  Conservative   61;  Mismatches  210;  Indels   86;  Gaps   10;

Qy       1 QTCPSVCSCS--NQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHL 58
Db      18 QICPKRCVCQILSPNLATLCAKKGLLFVPPNIDRRTVELRLADNFVTNIKRKDFANMTSL 77

Qy      59 EILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIES 118
Db      78 VDLTLSRNTISFITPHAFADLRNLRALHLNSNRLTKITNDMFSGLSNLHHLILNNNQLTL 137

Qy     119 IPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLD 178
Db     138 ISSTAFDDV----------------------------------------------FALE 150

Qy     179 ELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLL 238
Db     151 ELDLSYNNLETIPWDAVEKMVSLHTLSLDHNMIDNIPKGTFSHLHKMTRLDVTSNKLQKL 210
```

```
Qy     239 PHDLFTPLHHLERI--------------HLHHNPWNCNCDILWLSWWIKDMAPSNTACCA 284
           | |   ||    ::                    || :|||::|||         |
Db     211 PPD---PLFQRAQVLATSGIISPSTFALSFGGNPLHCNCELLWL------RRLSREDDLE 261

Qy     285 RCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRA-STSLTSVSWI 343
           | :|| | |||   : :  | |  |:|     :: | ||  | |:|:|      :: ||
Db     262 TCASPPLLTGRYFWSIPEEEFLCEPPLITRHTHEMRVLEGQRATLRCKARGDPEPAIHWI 321

Qy     344 TPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAAT 403
           :| | ::::     | |  :|||:    ||:||| :||: ||  |  |    |::
Db     322 SPEGKLISNA---TRSLVYDNGTLDILITTVKDTGAFTCIASNPAGEATQIVDLHI---I 375

Qy     404 TTPFSYFSTVTVETMEPSQDEARTTDNNVGPTPVVDWETTNVTTSLTPQSTRSTEKTFTI 463
             |   ||  :  :|   :  |:  :  |      ::|  | |:       | | :
Db     376 KLPHLLNSTNHIHEPDPGSSDISTSTKSGSNT-----SSSNGDTKLSQDKIVVAEATSST 430

Qy     464 PVTDIN--SGIPGI 475
           :   |    ||||
Db     431 ALLKFNFQRNIPGI 444


RESULT  11
Q9V430
ID   Q9V430      PRELIMINARY;      PRT;  1021 AA.
AC   Q9V430;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KEK3 PROTEIN.
GN   KEK3 OR BG:DS04862.1 OR CG4192.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Y, AND CN BW SP;
RX   MEDLINE=99403001; PubMed=10471707;
RA   Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,
RA   Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA   Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA   Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,
RA   Celniker S., Rubin G.M.;
RT   "An exploration of the sequence of a 2.9-Mb region of the genome of
RT   Drosophila melanogaster: the Adh region.";
RL   Genetics 153:179-219(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Y, AND CN BW SP;
RA   Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,
RA   Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA   Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA   Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,
RA   Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA   Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,
RA   Sethi H., Snir E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
RA   Zieran L.L., Rubin G.M.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003648; AAF53467.1; -.
DR   EMBL; AE003413; AAF44948.1; -.
DR   FlyBase; FBgn0028370; kek3.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 6.
DR   Pfam; PF01463; LRRCT; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00370; LRR; 2.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 2.
KW   Immunoglobulin domain.
SQ   SEQUENCE   1021 AA;  109770 MW;  23CB5712EFD60969 CRC64;


  Query Match             17.6%;  Score 443.5;  DB 5;  Length 1021;
  Best Local Similarity   30.0%;  Pred. No. 7.8e-25;
  Matches  127;  Conservative   54;  Mismatches  144;  Indels   99;  Gaps   13;

Qy       3 CPSVCSCSNQFSK--VICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSF--KHLRHL 58
           ||:|| |  :  |  |:|:  ||  :|  :   |:||:|  |:||:|  :||    | :|
Db      79 CPAVCECKWKSGKESVLCLNANLTHIPQPLDAGTQLLDLSGNEIQLIPDDSFATAQLLNL 138

Qy      59 EILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIES 118
             : : |:| |:| ||  ||  | ||  |:|  | |: ||: |  ::|:|:|| |  |||
Db     139 QKVYLARCHLRLIERHAFRKLINLVELDLSQNLLSAIPSLALYHVSELRELRLSGNPILR 198

Qy     119 IPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLD 178
           :|  ||  :| | :|:| :  |||:|:  || || :                    |:
Db     199 VPDDAFGHVPQLVKLELSDC-RLSHIAVRAFAGLES--------------------SLE 236

Qy     179 ELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLL 238
           | | || || :| |:   |
Db     237 WLKLDGNRLSEVRSGTITSL---------------------------------------- 256

Qy     239 PHDLFTPLHHLERIHLHHNPWNCNCDILWL-SWWIKDMAPSNTACCARCNTPPNLKGRYI 297
             || ||   |  | |||:| :  | :| ::   ||       | :|| | ||
Db     257 -----ASLHGLE---LARNTWNCSCSLRPLRAWMLQQNIPSGIP--PTCESPPRLSGRAW 306
```

```
Qy    298 GELDQNYFTCYAPVIVEPPADLNVTEGMAAELKCRA-STSLTSVSWITPNGTVMTHGA-- 354
          :|| : | |  | ||      :  ||    : |        :| |:  |  :    |
Db    307 DKLDVDDFAC-VPQIVATDTTAHGVEGRNITMSCYVEGVPQPAVKWLLKNRLIANLSAGG 365

Qy    355 -----YKVRIAVLSDG--------TLNFTNVTV-----QDTGMYTCMVSNSVGNTTASAT 396
               : | |  : |         | :|:|:     || |:|||   |  |   || |
Db    366 DGDSDSEPRTAAATQGRKTYVVNMLRNASNLTILTADMQDAGIYTCAAENKAGKVEASVT 425

Qy    397 LNVT 400
          | |:
Db    426 LAVS 429


RESULT  12
O43377
ID   O43377      PRELIMINARY;      PRT;   705 AA.
AC   O43377;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PROBABLE LEUCINE-RICH REPEAT PROTEIN.
GN   RG118D07.1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FIBROBLAST;
RA   Murray J., Langston Y., Clarke C.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC004142; AAC02752.1; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00041; fn3; 1.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 9.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00370; LRR; 4.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   705 AA;  79049 MW;  9B46066693F4BC1A CRC64;


  Query Match           16.7%;  Score 421;  DB 4;  Length 705;
  Best Local Similarity   27.1%;  Pred. No. 2.3e-23;
  Matches  133;  Conservative   74;  Mismatches  185;  Indels   98;  Gaps   12;

Qy      3 CPSVCSCS-----------NQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIK--- 48
            || :|:|           : | | |    |   |  :  ||::| |  | |  |:
Db     26 CPRLCTCEIRPWFTPRSIYMEASTVDCNDLGLLTFPARLPANTQILLLQTNNIAKIEYST 85

Qy     49 ---VN-----------------SFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELF 88
             ||                 : | :  |  : |  | :  :     : |:||  | :
Db     86 DFPVNLTGLDLSQNNLSSVTNINVKKMPQLLSVYLEENKLTELPEKCLSELSNLQELYIN 145

Qy     89 DNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGA 148
           | |:||  |||: |  |  | | :| :: | |  |: :|:|  | :||   :  | :
Db    146 HNLLSTISPGAFIGLHNLLRLHLNSNRLQMINSKWFDALPNLEILMIGE-NPIIRIKDMN 204

Qy    149 FEGLSNLRYLNLAMCNLREIPN-------------------------LTPLIKLDELDL 182
```

```
          |: | ||| | :|  || |||:                          |  :: |  |||
Db    205 FKPLINLRSLVIAGINLTEIPDNALVGLENLESISFYDNRLIKVPHVALQKVVNLKFLDL 264

Qy    183 SGNHLSAIRPGSFQGLMHLQKLWM--------IQS-----------------QIQVIER 216
          : | :: || | |  ::||::| :         | |                 ::  |
Db    265 NKNPINRIRRGDFSNMLHLKELGINNMPELISIDSLAVDNLPDLRKIEATNNPRLSYIHP 324

Qy    217 NAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSW---WIK 273
          |||  |  |  : |  | |: | |     | :|: | :| ||  |:| | |::     |:
Db    325 NAFFRLPKLESLMLNSNALSALYHGTIESLPNLKEISIHSNPIRCDCVIRWMNMNKTNIR 384

Qy    274 DMAPSNTACCARCNTPPNLKGRYIGELD-QNYFTCYAPVIVEP--PADLNVTEGMAAELK 330
            | | :  |      ||  :|: : ::  ::       |:|     |::|||  |
Db    385 FMEPDSLFCV----DPPEFQGQNVRQVHFRDMMEICLPLIAPESFPSNLNVEAGSYVSFH 440

Qy    331 CRASTS-LTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVG 389
          |||:      : ||||:|  :       :  | |:|||:   || :: |:|||: :| ||
Db    441 CRATAEPQPEIYWITPSGQKLLPNTLTDKFYVHSEGTLDINGVTPKEGGLYTCIATNLVG 500

Qy    390 NTTASATLNV 399
            |  : |
Db    501 ADLKSVMIKV 510


RESULT  13
Q9H3W5
ID   Q9H3W5      PRELIMINARY;      PRT;   708 AA.
AC   Q9H3W5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 79.4 KDA PROTEIN (NEURONAL LEUCINE-RICH REPEAT
DE   PROTEIN-3).
GN   DKFZP761K2424 OR NLRR-3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=AMYGDALA;
RA   Bloecker H., Boecher M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Hamano S., Inuzuka H., Morohashi A., Ohira M., Nakagawara A.;
RT   "Human neuronal leucine-rich repeat protein-3(NLRR-3).";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL442092; CAC09450.1; -.
DR   EMBL; AB060967; BAB47184.1; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00041; fn3; 1.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00560; LRR; 9.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00060; FN3; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00370; LRR; 4.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 8.
KW   Hypothetical protein; Immunoglobulin domain.
```

```
SQ   SEQUENCE   708 AA;  79424 MW;  24710478D6B124D1 CRC64;

  Query Match             16.7%;  Score 421;  DB 4;  Length 708;
  Best Local Similarity   27.1%;  Pred. No. 2.3e-23;
  Matches  133;  Conservative   74;  Mismatches  185;  Indels   98;  Gaps   12;

Qy       3 CPSVCSCS-----------NQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIK--- 48
           || :|:|             : | | |    |   |  :  ||::| |  | |  |:
Db      29 CPRLCTCEIRPWFTPRSIYMEASTVDCNDLGLLTFPARLPANTQILLLQTNNIAKIEYST 88

Qy      49 ---VN----------------SFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELF 88
             ||                 : | :  |  : |  | :  :      :|:||  | :
Db      89 DFPVNLTGLDLSQNNLSSVTNINVKKMPQLLSVYLEENKLTELPEKCLSELSNLQELYIN 148

Qy      89 DNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGA 148
           | |:||  |||: |  |  | | :| :: | |  |: :|:|  | :||   :  | :
Db     149 HNLLSTISPGAFIGLHNLLRLHLNSNRLQMINSKWFDALPNLEILMIGE-NPIIRIKDMN 207

Qy     149 FEGLSNLRYLNLAMCNLREIPN-------------------------LTPLIKLDELDL 182
           |: | ||| | :|  || |||:                         |  :: |  |||
Db     208 FKPLINLRSLVIAGINLTEIPDNALVGLENLESISFYDNRLIKVPHVALQKVVNLKFLDL 267

Qy     183 SGNHLSAIRPGSFQGLMHLQKLWM--------IQS-----------------QIQVIER 216
           : | :: || | |  ::||::| :        | |                 ::  |
Db     268 NKNPINRIRRGDFSNMLHLKELGINNMPELISIDSLAVDNLPDLRKIEATNNPRLSYIHP 327

Qy     217 NAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSW---WIK 273
           |||  |  |  : |  | |: | |     | :|: | :| ||  |:| | |::     |:
Db     328 NAFFRLPKLESLMLNSNALSALYHGTIESLPNLKEISIHSNPIRCDCVIRWMNMNKTNIR 387

Qy     274 DMAPSNTACCARCNTPPNLKGRYIGELD-QNYFTCYAPVIVEP--PADLNVTEGMAAELK 330
           | | :  |      ||  :|: : ::  ::      |:|     |::|||  |
Db     388 FMEPDSLFCV----DPPEFQGQNVRQVHFRDMMEICLPLIAPESFPSNLNVEAGSYVSFH 443

Qy     331 CRASTS-LTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVG 389
           |||:      : ||||:|  :       :  | |:|||:   || :: |:|||: :| ||
Db     444 CRATAEPQPEIYWITPSGQKLLPNTLTDKFYVHSEGTLDINGVTPKEGGLYTCIATNLVG 503

Qy     390 NTTASATLNV 399
               |  : |
Db     504 ADLKSVMIKV 513


RESULT  14
Q96JA1
ID   Q96JA1      PRELIMINARY;      PRT;  1093 AA.
AC   Q96JA1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MEMBRANE GLYCOPROTEIN LIG-1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Nilsson J., Vallbo C., Henriksson R., Hedman H.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF381545; AAK62357.1; -.
SQ   SEQUENCE   1093 AA;  119052 MW;  8488FA41BB10F353 CRC64;

  Query Match             16.4%;  Score 413.5;  DB 4;  Length 1093;
  Best Local Similarity   28.3%;  Pred. No. 1.5e-22;
  Matches  127;  Conservative   78;  Mismatches  170;  Indels   73;  Gaps   15;

Qy      37 LNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIP 96
           |:|: |:|::|:  :|: |  ||:|:| ||:|  :  ||| ||: :: | |  | |  :
Db     216 LDLNRNRIRLIEGLTFQGLNSLEVLKLQRNNISKLTDGAFWGLSKMHVLHLEYNSLVEVN 275

Qy      97 NGAFVYLSKLKELWLRNNPIESIPS------------YAFNRIPSLRRLDLGELKRL-- 141
           :|:   |: | :| | || |  |                :|| :  |    | ||  |
Db     276 SGSLYGLTALHQLHLSNNSIARIHRKGWSFCQKLHELVLSFNNLTRLDEESLAELSSLSV 335

Qy     142 --------SYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPG 193
                   |:|:||||:|| :|| |:                  ||  ::||        |
Db     336 LRLSHNSISHIAEGAFKGLRSLRVLD----------------LDHNEISGTIEDT--SG 376

Qy     194 SFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIH 253
           :| ||  | || :  ::|: : : ||  |: |  :||  | :  :  | |  : :|: :|
Db     377 AFSGLDSLSKLTLFGNKIKSVAKRAFSGLEGLEHLNLGGNAIRSVQFDAFVKMKNLKELH 436

Qy     254 LHHNPWNCNCDILWLSWWI--KDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTC---Y 308
           :  : : |:| : ||  |:  : :    || ||    | :|||: |  :    | |
Db     437 ISSDSFLCDCQLKWLPPWLIGRMLQAFVTATCAH---PESLKGQSIFSVPPESFVCDDFL 493

Qy     309 APVIVEPPADLNVTEGMAAELKCRASTSLTS---VSWITPNGTVMTHGAYKVRIAV-LSD 364
           | |:  |       |      | |::| :|    :|   |  |:|:   :  : |   |
Db     494 KPQIITQPETTMAMVGKDIRFTCSAASSSSSPMTFAWKKDN-EVLTNADMENFVHVHAQD 552

Qy     365 G-------TLNFTNVTVQDTGMYTCMVSNSVGNT---TASATLNVTAA-TTTPFSYFSTV 413
           |       |:   ||    | | |:::|  }:|    |  |:||  : | ||      :
Db     553 GEVMEYTTILHLRQVTFGHEGRYQCVITNHFGSTYSHKARLTVNVLPSFTKTPHD----I 608

Qy     414 TVETMEPSQDEARTTDNNVGPTPVVDWE 441
           |: |   :: |   | :   | | : |:
Db     609 TIRTTTVARLECAATGH---PNPQIAWQ 633


RESULT  15
Q9BYB8
ID   Q9BYB8      PRELIMINARY;      PRT;  1094 AA.
AC   Q9BYB8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MEMBRANE GLYCOPROTEIN LIG-1.
GN   LIG-1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Suzuki Y.;
RT   "human membrane glycoprotein LIG-1.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB050468; BAB40659.1; -.
DR   HSSP; P09661; 1A9N.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003591; LRR_typ.
DR   Pfam; PF00047; ig; 3.
DR   Pfam; PF00560; LRR; 15.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF01462; LRRNT; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00409; IG; 3.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00370; LRR; 5.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 14.
KW   Immunoglobulin domain.
SQ   SEQUENCE   1094 AA;  119165 MW;  6B2D0CC3C2783F18 CRC64;
```

```
  Query Match             16.4%;  Score 413.5;  DB 4;  Length 1094;
  Best Local Similarity   28.3%;  Pred. No. 1.5e-22;
  Matches  127;  Conservative   78;  Mismatches  170;  Indels   73;  Gaps   15;

Qy      37 LNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIP 96
           |:|: |:|::|:  :|: |  ||:|:| ||:|  :  ||| ||: :: | |  | |  :
Db     216 LDLNRNRIRLIEGLTFQGLNSLEVLKLQRNNISKLTDGAFWGLSKMHVLHLEYNSLVEVN 275

Qy      97 NGAFVYLSKLKELWLRNNPIESIPS------------YAFNRIPSLRRLDLGELKRL-- 141
           :|:   |: | :| | || |  |                :|| :  |    | ||  |
Db     276 SGSLYGLTALHQLHLSNNSIARIHRKGWSFCQKLHELVLSFNNLTRLDEESLAELSSLSV 335

Qy     142 --------SYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPG 193
                   |:|:||||:|| :|| |:                  ||  ::||        |
Db     336 LRLSHNSISHIAEGAFKGLRSLRVLD----------------LDHNEISGTIEDT--SG 376

Qy     194 SFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIH 253
           :| ||  | || :  ::|: : : ||  |: |  :||  | :  :  | |  : :|: :|
Db     377 AFSGLDSLSKLTLFGNKIKSVAKRAFSGLEGLEHLNLGGNAIRSVQFDAFVKMKNLKELH 436

Qy     254 LHHNPWNCNCDILWLSWWI--KDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTC---Y 308
           :  : : |:| : ||  |:  : :    || ||    | :|||: |  :    | |
Db     437 ISSDSFLCDCQLKWLPPWLIGRMLQAFVTATCAH---PESLKGQSIFSVPPESFVCDDFL 493

Qy     309 APVIVEPPADLNVTEGMAAELKCRASTSLTS---VSWITPNGTVMTHGAYKVRIAV-LSD 364
            | |:  |       |      | |::| :|    :|   |  |:|:   :  : |   |
Db     494 KPQIITQPETTMAMVGKDIRFTCSAASSSSSPMTFAWKKDN-EVLTNADMENFVHVHAQD 552

Qy     365 G-------TLNFTNVTVQDTGMYTCMVSNSVGNT---TASATLNVTAA-TTTPFSYFSTV 413
           |       |:   ||    | | |:::|  |:|    |  |:||  : | ||      :
Db     553 GEVMEYTTILHLRQVTFGHEGRYQCVITNHFGSTYSHKARLTVNVLPSFTKTPHD----I 608

Qy     414 TVETMEPSQDEARTTDNNVGPTPVVDWE 441
           |: |   :: |   | :   | | : |:
Db     609 TIRTTTMARLECAATGH---PNPQIAWQ 633


Search completed: August 26, 2002, 15:50:24
Job time: 246 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 2002, 15:43:48 ; Search time 16.37 Seconds
(without alignments)
713.221 Million cell updates/sec

Title: US-09-905-056-292_COPY_45_522
Perfect score: 2524
Sequence: 1 QTCPSVCSCSNQFSKVICVR.....KTFTIPVTDINSGIPGIDEV 478

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	421	16.7	708	4	US-09-131-648-2	Sequence 2, Appli
2	402.5	15.9	1091	3	US-08-986-485-5	Sequence 5, Appli
3	393	15.6	1101	3	US-08-986-485-2	Sequence 2, Appli
4	386	15.3	1525	3	US-09-191-647-2	Sequence 2, Appli
5	386	15.3	1525	4	US-09-540-245A-2	Sequence 2, Appli
6	386	15.3	1525	4	US-09-540-153-2	Sequence 2, Appli
7	372	14.7	1523	4	US-09-182-024A-2	Sequence 2, Appli
8	362	14.3	673	4	US-09-063-950-2	Sequence 2, Appli
9	354	14.0	649	4	US-09-188-930-305	Sequence 305, App
10	350	13.9	1480	3	US-09-191-647-7	Sequence 7, Appli
11	350	13.9	1480	4	US-09-540-245A-7	Sequence 7, Appli
12	350	13.9	1480	4	US-09-540-153-7	Sequence 7, Appli
13	350	13.9	1480	5	PCT-US91-09055-2	Sequence 2, Appli
14	347	13.7	1480	4	US-09-182-024A-5	Sequence 5, Appli
15	329.5	13.1	560	3	US-08-592-500-2	Sequence 2, Appli
16	329.5	13.1	560	3	US-08-195-006-2	Sequence 2, Appli
17	329.5	13.1	560	4	US-09-063-950-4	Sequence 4, Appli
18	329.5	13.1	560	5	PCT-US94-07644A-2	Sequence 2, Appli
19	329.5	13.1	605	4	US-09-063-950-5	Sequence 5, Appli
20	314.5	12.5	605	1	US-08-190-802A-49	Sequence 49, Appl
21	314.5	12.5	605	4	US-08-477-346-49	Sequence 49, Appl
22	314.5	12.5	605	4	US-08-473-089-49	Sequence 49, Appl
23	309	12.2	603	1	US-08-190-802A-50	Sequence 50, Appl
24	309	12.2	603	4	US-08-477-346-50	Sequence 50, Appl
25	309	12.2	603	4	US-08-473-089-50	Sequence 50, Appl
26	303.5	12.0	353	3	US-08-986-485-6	Sequence 6, Appli
27	302	12.0	282	1	US-08-442-063A-45	Sequence 45, Appl
28	302	12.0	307	1	US-08-442-063A-48	Sequence 48, Appl
29	302	12.0	333	1	US-08-442-063A-27	Sequence 27, Appl
30	302	12.0	342	1	US-08-272-919-2	Sequence 2, Appli
31	302	12.0	342	1	US-08-619-916-2	Sequence 2, Appli
32	302	12.0	342	5	PCT-US95-08542-2	Sequence 2, Appli
33	302	12.0	359	1	US-08-303-238-4	Sequence 4, Appli
34	302	12.0	359	4	US-08-458-834-4	Sequence 4, Appli
35	295	11.7	353	6	5340934-4	Patent No. 5340934
36	293	11.6	222	5	PCT-US91-09055-3	Sequence 3, Appli
37	293	11.6	231	3	US-08-986-485-7	Sequence 7, Appli
38	285	11.3	610	1	US-07-821-717B-6	Sequence 6, Appli
39	285	11.3	610	1	US-08-119-262B-6	Sequence 6, Appli
40	285	11.3	610	1	US-08-135-929A-11	Sequence 11, Appl
41	285	11.3	610	1	US-08-234-265A-11	Sequence 11, Appl
42	276	10.9	440	3	US-08-985-335-3	Sequence 3, Appli
43	276	10.9	440	4	US-09-410-372-3	Sequence 3, Appli
44	275.5	10.9	368	1	US-08-303-238-3	Sequence 3, Appli
45	275.5	10.9	368	4	US-08-458-834-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-131-648-2
; Sequence 2, Application US/09131648
; Patent No. 6168920
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: EXTRACELLULAR ADHESIVE PROTEINS
; FILE REFERENCE: PF-0576 US
; CURRENT APPLICATION NUMBER: US/09/131,648
; CURRENT FILING DATE: 1998-08-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 2687731
US-09-131-648-2

Query Match 16.7%; Score 421; DB 4; Length 708;
Best Local Similarity 27.1%; Pred. No. 7.9e-27;
Matches 133; Conservative 74; Mismatches 185; Indels 98; Gaps 12;

```
Qy       3 CPSVCSCS----------NQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIK--- 48
           || :|:|             : | | |    |   |  :  ||::| |  | |  |:
Db      29 CPRLCTCEIRPWFTPRSIYMEASTVDCNDLGLLTFPARLPANTQILLLQTNNIAKIEYST 88

Qy      49 ---VN----------------SFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELF 88
             ||                 : | :  |  :|  | :  :     : |:||  | :
Db      89 DFPVNLTGLDLSQNNLSSVTNINVKKMPQLLSVYLEENKLTELPEKCLSELSNLQELYIN 148

Qy      89 DNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGA 148
           | |:||  |||: |  |  | | :| :: | |  |: :|:|  | :||   :  | :
Db     149 HNLLSTISPGAFIGLHNLLRLHLNSNRLQMINSKWFDALPNLEILMIGE-NPIIRIKDMN 207

Qy     149 FEGLSNLRYLNLAMCNLREIPN-------------------------LTPLIKLDELDL 182
           |: | ||| | :|  || |||:                         |  :: |  |||
Db     208 FKPLINLRSLVIAGINLTEIPDNALVGLENLESISFYDNRLIKVPHVALQKVVNLKFLDL 267

Qy     183 SGNHLSAIRPGSFQGLMHLQKLWM--------IQS-----------------QIQVIER 216
           : | :: || | |  ::||::| :        | |                 ::  |
Db     268 NKNPINRIRRGDFSNMLHLKELGINNMPELISIDSLAVDNLPDLRKIEATNNPRLSYIHP 327

Qy     217 NAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSW---WIK 273
```

```
Db    328 NAFFRLPKLESLMLNSNALSALYHGTIESLPNLKEISIHSNPIRCDCVIRWMNMNKTNIR 387

Qy    274 DMAPSNTACCARCNTPPNLKGRYIGELD-QNYFTCYAPVIVEP--PADLNVTEGMAAELK 330
Db    388 FMEPDSLFCV----DPPEFQGQNVRQVHFRDMMEICLPLIAPESFPSNLNVEAGSYVSFH 443

Qy    331 CRASTS-LTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQDTGMYTCMVSNSVG 389
Db    444 CRATAEPQPEIYWITPSGQKLLPNTLTDKFYVHSEGTLDINGVTPKEGGLYTCIATNLVG 503

Qy    390 NTTASATLNV 399
Db    504 ADLKSVMIKV 513


RESULT   2
US-08-986-485-5
; Sequence 5, Application US/08986485
; Patent No. 6046030
;  GENERAL INFORMATION:
;    APPLICANT:  WU, SHUJIAN
;    APPLICANT:  SWEET, RAYMOND
;    APPLICANT:  TRUNEH, ALEMSEGED
;    TITLE OF INVENTION:  A HUMAN LIG-1 HOMOLOG (HLIG-1)
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RATNER & PRESTIA
;      STREET:  P.O. BOX 980
;      CITY:  VALLEY FORGE
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/986,485
;      FILING DATE:  08-DEC-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/059,448
;      FILING DATE:  22-SEP-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PRESTIA, PAUL F
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  GH-70264
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1091 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-986-485-5


  Query Match             15.9%;  Score 402.5;  DB 3;  Length 1091;
  Best Local Similarity   24.1%;  Pred. No. 5.1e-25;
  Matches  134;  Conservative   77;  Mismatches  168;  Indels  177;  Gaps   14;

Qy     37 LNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIP 96
Db    218 LDLNRNRIRLIEGLTFQGLDSLEVLRLQRNNISRLTDGAFWGLSKMHVLHLEYNSLVEVN 277

Qy     97 NGAFVYLSKLKELWLRNNPIESIPS------------YAFNRIPSLRRLDLGELKRL-- 141
```

```
Db    278 SGSLYGLTALHQLHLSNNSISRIQRDGWSFCQKLHELILSFNNLTRLDEESLAELSSLSI 337

Qy    142 --------SYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPG 193
Db    338 LRLSHNAISHIAEGAFKGLKSLRVLD----------------LDHNEISGTIEDT--SG 378

Qy    194 SFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIH 253
Db    379 AFTGLDNLSKLTLFGNKIKSVAKRAFSGLESLEHLNLGENAIRSVQFDAFAKMKNLKELY 438

Qy    254 LHHNPWNCNCDILWLSWWI--KDMAPSNTACCARCNTPPNLKGRYI-GELDQNY------ 304
Db    439 ISSESFLCDCQLKWLPPWLMGRMLQAFVTATCAH---PESLKGQSIFSVLPDSFVCDDFP 495

Qy    305 --------------------FTCYA---------------------------------- 309
Db    496 KPQIITQPETTMAVVGKDIRFTCSAASSSSSPMTFAWKKDNEVLANADMENFAHVRAQDG 555

Qy    310 -------------------------------------------PVIVEPPADLNVTEG 324
Db    556 EVMEYTTILHLRHVTFGHEGRYQCIITNHFGSTYSHKARLTVNVLPSFTKIPHDIAIRTG 615

Qy    325 MAAELKCRASTSLT-SVSWITPNGTVMTHGAYKVRIAVLSDGTLNF-TNVTVQDTGMYTC 382
Db    616 TTARLECAATGHPNPQIAWQKDGGTDFP-AARERRMHVMPDDDVFFITDVKIDDMGVYSC 674

Qy    383 MVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNV--------GP 434
Db    675 TAQNSAGSVSANATLTVLE-----------TPSLAVPLEDRVVTVGETVAFQCKATGSP 722

Qy    435 TPVVDWETTNVTTSLT 450
Db    723 TPRITWLKGGRPLSLT 738


RESULT   3
US-08-986-485-2
; Sequence 2, Application US/08986485
; Patent No. 6046030
;  GENERAL INFORMATION:
;    APPLICANT:  WU, SHUJIAN
;    APPLICANT:  SWEET, RAYMOND
;    APPLICANT:  TRUNEH, ALEMSEGED
;    TITLE OF INVENTION:  A HUMAN LIG-1 HOMOLOG (HLIG-1)
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RATNER & PRESTIA
;      STREET:  P.O. BOX 980
;      CITY:  VALLEY FORGE
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/986,485
;      FILING DATE:  08-DEC-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/059,448
;      FILING DATE:  22-SEP-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PRESTIA, PAUL F
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  GH-70264
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0701
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  2:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1101 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-986-485-2


  Query Match             15.6%;  Score 393;  DB 3;  Length 1101;
  Best Local Similarity   27.8%;  Pred. No. 3.2e-24;
  Matches  125;  Conservative   79;  Mismatches  171;  Indels   74;  Gaps   16;

Qy      37 LNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIP 96
           |:|: |:|::|:  :|: |  ||:|:| ||:|  :  ||| ||: :: | |  : |  :
Db     223 LDLNRNRIRLIEGLTFQGLNSLEVLKLQRNNISKLTDGAFWGLSKMHVLHLEYDSLVEVN 282

Qy      97 NGAFVYLSKLKELWLRNNPIESIPS-------------YAFNRIPSLRRLDLGELKRL-- 141
           :|:   |: | :| | || |  |                  :|| :  |    | ||  |
Db     283 SGSLYGLTALHQLHLSNNSIARIHRKGWSFCQKLHELVLSFNNLTRLDEESLAELSSLSV 342

Qy     142 --------SYISEGAFEGLSNLRYLNLAMCNLREIPNLTPLIKLDELDLSGNHLSAIRPG 193
                  |:|:||||:|| :|| |:                 ||  ::||         |
Db     343 LRLSHNSISHIAEGAFKGLRSLRVLD-----------------LDHNEISGTIEDT--SG 383

Qy     194 SFQGL-MHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERI 252
           :| ||     || :  ::|: : : ||  |: |  :||  | :  :  | |  : :|: :
Db     384 AFSGLEFGHSKLTLFGNKIKSVAKRAFSGLEGLEHLNLGGNAIRSVQFDAFVKMKNLKEL 443

Qy     253 HLHHNPWNCNCDILWLSWWI--KDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTC--- 307
           |:  : : |:| : ||  |:  : :   || ||    | :|||: |  :    | |
Db     444 HISSDSFLCDCQLKWLPPWLIGRMLQAFVTATCAH---PESLKGQSIFSVPPESFVCDDF 500

Qy     308 YAPVIVEPPADLNVTEGMAAELKCRASTSLTS---VSWITPNGTVMTHGAYKVRIAV-LS 363
             | |:  |       |      | |::| :|    :|   |  |:|:   :  : |
Db     501 LKPQIITQPETTMAMVGKDIRFTCSAASSSSSPMTFAWKKDN-EVLTNADMENFVHVHAQ 559

Qy     364 DG-------TLNFTNVTVQDTGMYTCMVSNSVGNT---TASATLNVTAA-TTTPFSYFST 412
           ||        |:   ||    | | |:::|  |:|    |  |:||  : | ||
Db     560 DGEVMEYTTILHLRQVTFGHEGRYQCVITNHFGSTYSHKARLTVNVLPSFTKTPHD---- 615

Qy     413 VTVETMEPSQDEARTTDNNVGPTPVVDWE 441
           :|: |   :: |   | :   | | : |:
Db     616 ITIRTTTVARLECAATGH---PNPQIAWQ 641


RESULT   4
US-09-191-647-2
; Sequence 2, Application US/09191647
; Patent No. 6046015
; GENERAL INFORMATION:
;  APPLICANT: Goodman, Corey
;  APPLICANT: Kid, Thomas
;  APPLICANT: Brose, Katja
;  APPLICANT: Tessier-Lavigne, Marc
;  TITLE OF INVENTION: Modulating Robo: Ligand Interactions
;  FILE REFERENCE: B98-031-3
;  CURRENT APPLICATION NUMBER: US/09/191,647
;  CURRENT FILING DATE: 1998-11-13
;  EARLIER APPLICATION NUMBER: 60/065,544
;  EARLIER FILING DATE: 1997-11-14
;  EARLIER APPLICATION NUMBER: 60/081,057
;  EARLIER FILING DATE: 1998-04-07
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1525
;   TYPE: PRT
;   ORGANISM: human
US-09-191-647-2


  Query Match             15.3%;  Score 386;  DB 3;  Length 1525;
  Best Local Similarity   25.0%;  Pred. No. 2e-23;
  Matches  124;  Conservative   48;  Mismatches  164;  Indels  160;  Gaps    8;

Qy       1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
           | ||: ||||   | | |    || ||  |  ||  |:|: | |  |    |  |||| :
Db      26 QACPAQCSCSG--STVDCHGLALRSVPRNIPRNTERLDLNGNNITRITKTDFAGLRHLRV 83

Qy      61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
           |||  | | ||| |||  |  |  | |  | |    |   |:  :|| | |  | |::||
Db      84 LQLMENKISTIERGAFQDLKELERLRLNRNHLQLFPELLFLGTAKLYRLDLSENQIQAIP 143

Qy     121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNL---------------------- 160
             ||     :: | | :  ::| | :|||  | :|  | |
Db     144 RKAFRGAVDIKNLQL-DYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPKLR 202

Qy     161 ------------------------------------------------------------ 160

Db     203 TFRLHSNNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPSHLRGHNVAEVQKREFVCSDEEE 262

Qy     161 ----------------AMC------------NLREIPN---------------------- 170
                            | |             | |||
Db     263 GHQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPTNLPETITEIRLEQNTIKVIPPG 322

Qy     171 -LTPLIKLDELDLSGNHLSAIRPGSFQ-----------------------GLMHLQKLW 205
            :|  ||  :||| | :| : | :||                       ||  || |
Db     323 AFSPYKKLRRIDLSNNQISELAPDAFQGLRSLNSLVLYGNKITELPKSLFEGLFSLQLLL 382

Qy     206 MIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDI 265
           :  ::|  :  :|| :| :|  ::|  | |  :    |:||  :: :||  ||: |:| :
Db     383 LNANKINCLRVDAFQDLHNLNLLSLYDNKLQTIAKGTFSPLRAIQTMHLAQNPFICDCHL 442

Qy     266 LWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGM 325
           ||: ::    |  |:  ||| :|  |  : ||::    | |             :    :
Db     443 KWLADYL-HTNPIETS-GARCTSPRRLANKRIGQIKSKKFRCSGTEDYRSKLSGDCFADL 500

Qy     326 AAELKCRASTSLTSVS 341
           |   |||   :   |
Db     501 ACPEKCRCEGTTVDCS 516


RESULT   5
US-09-540-245A-2
; Sequence 2, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
;  APPLICANT:  Goodman, Corey
;  APPLICANT:  Kid, Thomas
;  APPLICANT:  Brose, Katja
;  APPLICANT:  Tessier-Lavigne, Marc
;  TITLE OF INVENTION:  Modulating Robo: Ligand Interactions
;  FILE REFERENCE:  B98-031-3
;  CURRENT APPLICATION NUMBER: US/09/540,245A
;  CURRENT FILING DATE:  2000-03-31
;  PRIOR APPLICATION NUMBER:  60/065,544
;  PRIOR FILING DATE:  1997-11-14
;  PRIOR APPLICATION NUMBER:  60/081,057
;  PRIOR FILING DATE:  1998-04-07
;  NUMBER OF SEQ ID NOS:  20
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1525
;   TYPE: PRT
;   ORGANISM: human
US-09-540-245A-2


  Query Match             15.3%;  Score 386;  DB 4;  Length 1525;
  Best Local Similarity   25.0%;  Pred. No. 2e-23;
  Matches  124;  Conservative   48;  Mismatches  164;  Indels  160;  Gaps    8;

Qy       1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
           | ||: ||||   | | |    || ||  |  ||  |:|: | |  |    |  |||| :
Db      26 QACPAQCSCSG--STVDCHGLALRSVPRNIPRNTERLDLNGNNITRITKTDFAGLRHLRV 83
```

```
Qy      61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
           |||  | | ||| |||  |  |  | |  | |    |   |:  :||  | |  | |::||
Db      84 LQLMENKISTIERGAFQDLKELERLRLNRNHLQLFPELLFLGTAKLYRLDLSENQIQAIP 143

Qy     121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNL-------------------- 160
           ||      :: | | :  ::| | :|||  | :|  | |
Db     144 RKAFRGAVDIKNLQL-DYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPKLR 202

Qy     161 ------------------------------------------------------------ 160

Db     203 TFRLHSNNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPSHLRGHNVAEVQKREFVCSDEEE 262

Qy     161 ----------------AMC------------NLREIPN---------------------- 170
                           | |              | |||
Db     263 GHQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPTNLPETITEIRLEQNTIKVIPPG 322

Qy     171 -LTPLIKLDELDLSGNHLSAIRPGSFQ-----------------------GLMHLQKLW 205
            :|  ||  :||| | :| : | :||                       ||  || |
Db     323 AFSPYKKLRRIDLSNNQISELAPDAFQGLRSLNSLVLYGNKITELPKSLFEGLFSLQLLL 382

Qy     206 MIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDI 265
           :  ::|  :  :|| :| :|  ::|  | |  :    |:||  :: :||  ||: |:| :
Db     383 LNANKINCLRVDAFQDLHNLNLLSLYDNKLQTIAKGTFSPLRAIQTMHLAQNPFICDCHL 442

Qy     266 LWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGM 325
           ||: ::    |  |:  ||| :|  |  : ||::    | |            :    :
Db     443 KWLADYL-HTNPIETS-GARCTSPRRLANKRIGQIKSKKFRCSGTEDYRSKLSGDCFADL 500

Qy     326 AAELKCRASTSLTSVS 341
           |   |||   :   |
Db     501 ACPEKCRCEGTTVDCS 516
```

RESULT 6
US-09-540-153-2
; Sequence 2, Application US/09540153
; Patent No. 6270995
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,153
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/191,647
; PRIOR FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: human
US-09-540-153-2

Query Match 15.3%; Score 386; DB 4; Length 1525;
Best Local Similarity 25.0%; Pred. No. 2e-23;
Matches 124; Conservative 48; Mismatches 164; Indels 160; Gaps 8;

```
Qy       1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
           | ||: ||||   | | |    || ||  |  ||  |:|: | |  |    |  ||||  :
Db      26 QACPAQCSCSG--STVDCHGLALRSVPRNIPRNTERLDLNGNNITRITKTDFAGLRHLRV 83

Qy      61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
           |||  | | ||| |||  |  |  | |  | |    |   |:  :||  | |  | |::||
Db      84 LQLMENKISTIERGAFQDLKELERLRLNRNHLQLFPELLFLGTAKLYRLDLSENQIQAIP 143

Qy     121 SYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNL-------------------- 160
           ||      :: | | :  ::| | :|||  | :|  | |
Db     144 RKAFRGAVDIKNLQL-DYNQISCIEDGAFRALRDLEVLTLNNNNITRLSVASFNHMPKLR 202

Qy     161 ------------------------------------------------------------ 160

Db     203 TFRLHSNNLYCDCHLAWLSDWLRKRPRVGLYTQCMGPSHLRGHNVAEVQKREFVCSDEEE 262

Qy     161 ----------------AMC------------NLREIPN---------------------- 170
                           | |              | |||
Db     263 GHQSFMAPSCSVLHCPAACTCSNNIVDCRGKGLTEIPTNLPETITEIRLEQNTIKVIPPG 322

Qy     171 -LTPLIKLDELDLSGNHLSAIRPGSFQ-----------------------GLMHLQKLW 205
            :|  ||  :||| | :| : | :||                       ||  || |
Db     323 AFSPYKKLRRIDLSNNQISELAPDAFQGLRSLNSLVLYGNKITELPKSLFEGLFSLQLLL 382

Qy     206 MIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDI 265
           :  ::|  :  :|| :| :|  ::|  | |  :    |:||  :: :||  ||: |:| :
Db     383 LNANKINCLRVDAFQDLHNLNLLSLYDNKLQTIAKGTFSPLRAIQTMHLAQNPFICDCHL 442

Qy     266 LWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGM 325
           ||: ::    |  |:  ||| :|  |  : ||::    | |            :    :
Db     443 KWLADYL-HTNPIETS-GARCTSPRRLANKRIGQIKSKKFRCSGTEDYRSKLSGDCFADL 500

Qy     326 AAELKCRASTSLTSVS 341
           |   |||   :   |
Db     501 ACPEKCRCEGTTVDCS 516
```

RESULT 7
US-09-182-024A-2
; Sequence 2, Application US/09182024A
; Patent No. 6342370
; GENERAL INFORMATION:
; APPLICANT: Connolly, Timothy
; APPLICANT: Rajput, Bhanu
; TITLE OF INVENTION: Human Slit Polypeptide and Polynucleotides Encoding
; TITLE OF INVENTION: Same
; FILE REFERENCE: 640100-271
; CURRENT APPLICATION NUMBER: US/09/182,024A
; CURRENT FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: 60/063,946
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/096,420
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1523
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-182-024A-2

Query Match 14.7%; Score 372; DB 4; Length 1523;
Best Local Similarity 24.6%; Pred. No. 2.9e-22;
Matches 132; Conservative 56; Mismatches 173; Indels 176; Gaps 13;

```
Qy       3 CPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQ 62
           ||: |:||   : | |    || || ||  |   |:|  | |  |    |  ||::| :|
Db      34 CPTKCTCS--AASVDCHGLGLRAVPRGIPRNAERLDLDRNNITRITKMDFAGLKNLRVLH 91

Qy      63 LSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPSY 122
           |  | :  || |||  |  |  | |  |:|  :|   |    ||  | |  | |: ||
Db      92 LEDNQVSVIERGAFQDLKQLERLRLNKNKLQVLPELLFQSTPKLTRLDLSENQIQGIPRK 151

Qy     123 AFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNL---------------------- 160
           ||  |  :: | | :   :| | :|||  | :|  | |
Db     152 AFRGITDVKNLQL-DNNHISCIEDGAFRALRDLEILTLNNNNISRILVTSFNHMPKIRTL 210

Qy     161 ------------------------------------------------------------ 160

Db     211 RLHSNHLYCDCHLAWLSDWLRQRRTVGQFTLCMAPVHLRGFNVADVQKKEYVCPAPHSEP 270
```

```
Qy    161 AMCN----------------------LREIP-NL----------------------TPL 174
Db    271 PSCNANSISCPSPCTCSNNIVDCRGKGLMEIPANLPEGIVEIRLEQNSIKAIPAGAFTQY 330

Qy    175 IKLDELDLSGNHLSAIRPGSFQ-----------------------GLMHLQKLWMIQSQ 210
Db    331 KKLKRIDISKNQISDIAPDAFQGLKSLTSLVLYGNKITEIAKGLFDGLVSLQLLLLNANK 390

Qy    211 IQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSW 270
Db    391 INCLRVNTFQDLQNLNLLSLYDNKLQTISKGLFAPLQSIQTLHLAQNPFVCDCHLKWLAD 450

Qy    271 WIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVEPPADLNVTEGMAAELK 330
Db    451 YLQD-NPIETS-GARCSSPRRLANKRISQIKSKKFRCSGSEDYRSRFSSECFMDLVCPEK 508

Qy    331 CRASTSLTSVSWITPNGTVMTHGAYK-VRI-----AVLSDGTLNFTNVTV-QDTGMY 380
Db    509 CRC------------EGTIVDCSNQKLVRIPSHLPEYVTDLRLNDNEVSVLEATGIF 553


RESULT   8
US-09-063-950-2
; Sequence 2, Application US/09063950C
; Patent No. 6225085
; GENERAL INFORMATION:
;  APPLICANT: Holtzman, Douglas A.
;  TITLE OF INVENTION: NOVEL LRSG PROTEIN AND NUCLEIC ACID MOLECULES AND USES
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: MEI-019
;  CURRENT APPLICATION NUMBER: US/09/063,950C
;  CURRENT FILING DATE: 1998-04-21
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 673
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-063-950-2


  Query Match             14.3%;  Score 362;  DB 4;  Length 673;
  Best Local Similarity   29.9%;  Pred. No. 6e-22;
  Matches  115;  Conservative   56;  Mismatches  150;  Indels   64;  Gaps   11;

Qy      1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEI 60
Db     22 QGCPSGCQCS-QPQTVFCTARQGTTVPRDVPPDTVGLYVFENGITMLDAGSFAGLPGLQL 80

Qy     61 LQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIP 120
Db     81 LDLSQNQIASLPSGVFQPLANLSNLDLTANRLHEITNETFRGLRRLERLYLGKNRIRHIQ 140

Qy    121 SYAFN--------------------RIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLN 159
Db    141 PGAFDTLDRLLELKLQDNELRALPPLRLPRLLLLDLSHNSLLAL--EPGILDTANVEALR 198

Qy    160 LAMCNLREIPN--LTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMI-QSQIQVIER 216
Db    199 LAGLGLQQLDEGLFSRLRNLHDLDVSDNQLERV-PPVIRGLRGLTRLRLAGNTRIAQLRP 257

Qy    217 NAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKD-- 274
Db    258 EDLAGLAALQELDVSNLSLQALPGDLSGLFPRLRLLAAARNPFNCVCPLSWFGPWVRESH 317

Qy    275 ---MAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYA----------PVIVEPPADLN 320
Db    318 VTLASPEET----RCHFPPKNAGRLLLELDYADFGCPATTTTATVPTTRPVVREPTA--- 370

Qy    321 VTEGMAAELKCRASTSLTSVSWITP 345
Db    371 ------------LSSSL-APTWLSP 382


RESULT   9
US-09-188-930-305
; Sequence 305, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT: Strachan, Lorna
;  APPLICANT: Sleeman, Matthew
;  APPLICANT: Onrust, Rene
;  APPLICANT: Murison, James Greg
;  TITLE OF INVENTION: Compositions Isolated From Skin Cells
;  TITLE OF INVENTION: and Methods For Their Use
;  FILE REFERENCE: 11000.1011c1
;  CURRENT APPLICATION NUMBER: US/09/188,930A
;  CURRENT FILING DATE: 1998-11-09
;  NUMBER OF SEQ ID NOS: 348
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 305
;   LENGTH: 649
;   TYPE: PRT
;   ORGANISM: Mouse
US-09-188-930-305


  Query Match             14.0%;  Score 354;  DB 4;  Length 649;
  Best Local Similarity   24.6%;  Pred. No. 2.7e-21;
  Matches  136;  Conservative   88;  Mismatches  202;  Indels  126;  Gaps   20;

Qy      1 QTCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNS-FKHL---- 55
Db     29 KSCPSVCRCDAGF--IYCNDRSLTSIPVGIPEDATTLYLQNNQINNVGIPSDLKNLLKVQ 86

Qy     56 -----------------RHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLT--TIP 96
Db     87 RIYLYHNSLDEFPTNLPKYVKELHLQENNIRTITYDSLSKIPYLEELHLDDNSVSAVSIE 146

Qy     97 NGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLR 156
Db    147 EGAFRDSNYLRLLFLSRNHLSTIPGGLPRTIEELRLDD----NRISTISSPSLHGLTSLK 202

Qy    157 YLNLAMCNLREIPNLTP-----LIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQI 211
Db    203 RLVLD-GNLLNNHGLGDKVFFNLVNLTELSLVRNSLTA-APVNLPG-TSLRKLYLQDNHI 259

Qy    212 QVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWW 271
Db    260 NRVPPNAFSYLRQLYRLDMSNNNLSNLPQGIFDDLDNITQLILRNNPWYCGCKMKWVRDW 319

Qy    272 IKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCY---------------------- 308
Db    320 LQSLPVKVNVRGLMCQAPEKVRGMAIKDLSAELFDCKDSGIVSTIQITTAIPNTAYPAQG 379

Qy    309 ---APVIVEPP-------ADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVR 358
Db    380 QWPAPVTKQPDIKNPKLIKDQRTTGSPSRK------TILITVKSVTPD---TIHISWRLA 430

Qy    359 IAVLSDGTLNFTNVTVQDTGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFST------ 412
Db    431 LPM--------------TALRLSWL--KLGHSPAFGSITETIVTGERSEYLVTALEPES 473

Qy    413 ---VTVETMEPSQ----DEARTTDNNVGPTPV-VDWETT-----NVTTSLTPQSTRSTEK 459
Db    474 PYRVCMVPMETSNLYLFDE----------TPVCIETQTAPLRMYNPTTTLNREQEKEPYK 523

Qy    460 TFTIPVTDINSG 471
Db    524 NPNLPLAAIIGG 535


RESULT  10
US-09-191-647-7
; Sequence 7, Application US/09191647
```

; Patent No. 6046015
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/191,647
; CURRENT FILING DATE: 1998-11-13
; EARLIER APPLICATION NUMBER: 60/065,544
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: 60/081,057
; EARLIER FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1480
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-191-647-7

Query Match 13.9%; Score 350; DB 3; Length 1480;
Best Local Similarity 23.9%; Pred. No. 1.9e-20;
Matches 109; Conservative 65; Mismatches 152; Indels 130; Gaps 9;

```
Qy       2 TCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEIL 61
           :||  | |::    | |  |:|  ||  :  :|  : | :| |  :   ||   | |  :
Db     294 SCPHPCRCADGI--VDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKSFSSFRRLRRI 351

Qy      62 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 121
            || |:|  |   | :||  | || |: |::  :|:| |  |  |: | |  | |  |
Db     352 DLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLLNANEISCIRK 411

Qy     122 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM------CNLR--------- 166
            ||  : ||  | | :   :  :: | |: : ::: ::||       ||||
Db     412 DAFRDLHSLSLLSLYD-NNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRWLADYLHKN 470

Qy     167 ------------------------------------------------------------ 166

Db     471 PIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKLSGECRMDSDCPAMCHCEGTTV 530

Qy     167 --------EIPNLTPL--------------IKLD----------ELDLSGNHLSAIRPGS 194
                  |||   ||              |  |          :|:|  | |: | | :
Db     531 DCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKRNQLTGIEPNA 590

Qy     195 FQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 254
           |:|  |:|:| : :::|: |    |  |  |  :||  | :: :    |  |: |  ::|
Db     591 FEGASHIQELQLGENKIKEISNKMFLGLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNL 650

Qy     255 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVE 314
             ||:|||| : | :  ::  : :  |  |||  |  ::   | :|  : | | :
Db     651 ASNPFNCNCHLAWFAECVRKKSLNGGA--ARCGAPSKVRDVQIKDLPHSEFKCSS----- 703

Qy     315 PPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVM 350
                 :||   :  |  | : |        |||:
Db     704 -----ENSEGCLGDGYCPPSCTCT--------GTVV 726
```

RESULT 11
US-09-540-245A-7
; Sequence 7, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,245A
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1480
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-540-245A-7

Query Match 13.9%; Score 350; DB 4; Length 1480;
Best Local Similarity 23.9%; Pred. No. 1.9e-20;
Matches 109; Conservative 65; Mismatches 152; Indels 130; Gaps 9;

```
Qy       2 TCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEIL 61
           :||  | |::    | |  |:|  ||  :  :|  : | :| |  :   ||   | |  :
Db     294 SCPHPCRCADGI--VDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKSFSSFRRLRRI 351

Qy      62 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 121
            || |:|  |   | :||  | || |: |::  :|:| |  |  |: | |  | |  |
Db     352 DLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLLNANEISCIRK 411

Qy     122 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM------CNLR--------- 166
            ||  : ||  | | :   :  :: | |: : ::: ::||       ||||
Db     412 DAFRDLHSLSLLSLYD-NNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRWLADYLHKN 470

Qy     167 ------------------------------------------------------------ 166

Db     471 PIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKLSGECRMDSDCPAMCHCEGTTV 530

Qy     167 --------EIPNLTPL--------------IKLD----------ELDLSGNHLSAIRPGS 194
                  |||   ||              |  |          :|:|  | |: | | :
Db     531 DCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKRNQLTGIEPNA 590

Qy     195 FQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 254
           |:|  |:|:| : :::|: |    |  |  |  :||  | :: :    |  |: |  ::|
Db     591 FEGASHIQELQLGENKIKEISNKMFLGLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNL 650

Qy     255 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVE 314
             ||:|||| : | :  ::  : :  |  |||  |  ::   | :|  : | | :
Db     651 ASNPFNCNCHLAWFAECVRKKSLNGGA--ARCGAPSKVRDVQIKDLPHSEFKCSS----- 703

Qy     315 PPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVM 350
                 :||   :  |  | : |        |||:
Db     704 -----ENSEGCLGDGYCPPSCTCT--------GTVV 726
```

RESULT 12
US-09-540-153-7
; Sequence 7, Application US/09540153
; Patent No. 6270995
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,153
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/191,647
; PRIOR FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1480
; TYPE: PRT
; ORGANISM: Drosophila melanogaster

US-09-540-153-7

Query Match 13.9%; Score 350; DB 4; Length 1480;
Best Local Similarity 23.9%; Pred. No. 1.9e-20;
Matches 109; Conservative 65; Mismatches 152; Indels 130; Gaps 9;

```
Qy      2 TCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEIL 61
          :||  | |::    | |  |:|  ||  :  :|  : | :| |  :   ||   | |  :
Db    294 SCPHPCRCADGI--VDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKSFSSFRRLRRI 351

Qy     62 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 121
           || |:|  |   | :||  | || |: |::  :|:| |  |  |: | |  | |  |
Db    352 DLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLLNANEISCIRK 411

Qy    122 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM------CNLR--------- 166
           ||  : ||  | | :   :  :: | |: : ::: ::||       ||||
Db    412 DAFRDLHSLSLLSLYD-NNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRWLADYLHKN 470

Qy    167 ------------------------------------------------------------ 166

Db    471 PIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKLSGECRMDSDCPAMCHCEGTTV 530

Qy    167 --------EIPNLTPL--------------IKLD----------ELDLSGNHLSAIRPGS 194
                  |||   ||               |  |          :|:|  | |: | | :
Db    531 DCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKRNQLTGIEPNA 590

Qy    195 FQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 254
          |:|  |:|:| : :::|: |     |  |  |  :||  | :: :     |  |: |  ::|
Db    591 FEGASHIQELQLGENKIKEISNKMFLGLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNL 650

Qy    255 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVE 314
            ||:|||| : | :  ::  : :  |  |||  |  ::   | :|  : | | :
Db    651 ASNPFNCNCHLAWFAECVRKKSLNGGA--ARCGAPSKVRDVQIKDLPHSEFKCSS----- 703

Qy    315 PPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVM 350
                :||   :  |  | : |        |||:
Db    704 -----ENSEGCLGDGYCPPSCTCT--------GTVV 726
```

RESULT 13
PCT-US91-09055-2
; Sequence 2, Application PC/TUS9109055
; GENERAL INFORMATION:
; APPLICANT: Rothberg, Jonathan Marc and Artavanis-Tsakonas, Spyridon
; TITLE OF INVENTION: Purified SLIT protein and Sequence Elements Thereof
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University
; ADDRESSEE: Office of Cooperative Research
; STREET: 246 Church Street
; STREET: Suite 401
; CITY: New Haven
; STATE: Connecuticut
; COUNTRY: USA
; ZIP: 06510
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch. 800 Kb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/09055
; FILING DATE: 19911127
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/624,135
; FILING DATE: 7-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Barth, Richard J.
; REGISTRATION NUMBER: 28,180
; REFERENCE/DOCKET NUMBER: 900964/RSB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 972-1400
; TELEFAX: (212) 370-1622
; TELEX: 236268
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1480 amino acids
; TYPE: AMINO ACIDS
; TOPOLOGY: Linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: signal sequence
; LOCATION: 1 to 36
; IDENTIFICATION METHOD: similarity to other signal
; OTHER INFORMATION: Directs Export
; NAME/KEY: Four Flank-LRR-Flank domains
; LOCATION: 37 to 910
; IDENTIFICATION METHOD: Array of Flank-LRR-Flank
; OTHER INFORMATION: mediates adhesive events
; NAME/KEY: Tandem EGF-like repeats
; LOCATION: 911 to 1150
; IDENTIFICATION METHOD: similarity to tandem EGF-like r
; OTHER INFORMATION: protein-protein interactions
; NAME/KEY: 7th EGF-like repeat
; LOCATION: 1353 to 1393
; IDENTIFICATION METHOD: similarity to epidermal growth f
; OTHER INFORMATION: Involvement in receptor-ligand
; NAME/KEY: Alternative splice segment
; LOCATION: 1394 to 1404
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: developmentally regulated
; NAME/KEY: COOH-terminal region
; LOCATION: 1405 to 1480
; IDENTIFICATION METHOD: experimental
PCT-US91-09055-2

Query Match 13.9%; Score 350; DB 5; Length 1480;
Best Local Similarity 23.9%; Pred. No. 1.9e-20;
Matches 109; Conservative 65; Mismatches 152; Indels 130; Gaps 9;

```
Qy      2 TCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEIL 61
          :||  | |::    | |  |:|  ||  :  :|  : | :| |  :   ||   | |  :
Db    294 SCPHPCRCADGI--VDCREKSLTSVPVTLPDDTTDVRLEQNFITELPPKSFSSFRRLRRI 351

Qy     62 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 121
           || |:|  |   | :||  | || |: |::  :|:| |  |  |: | |  | |  |
Db    352 DLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLLNANEISCIRK 411

Qy    122 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM------CNLR--------- 166
           ||  : ||  | | :   :  :: | |: : ::: ::||       ||||
Db    412 DAFRDLHSLSLLSLYD-NNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRWLADYLHKN 470

Qy    167 ------------------------------------------------------------ 166

Db    471 PIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKLSGECRMDSDCPAMCHCEGTTV 530

Qy    167 --------EIPNLTPL--------------IKLD----------ELDLSGNHLSAIRPGS 194
                  |||   ||               |  |          :|:|  | |: | | :
Db    531 DCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKRNQLTGIEPNA 590

Qy    195 FQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 254
          |:|  |:|:| : :::|: |     |  |  |  :||  | :: :     |  |: |  ::|
Db    591 FEGASHIQELQLGENKIKEISNKMFLGLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNL 650

Qy    255 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVE 314
            ||:|||| : | :  ::  : :  |  |||  |  ::   | :|  : | | :
Db    651 ASNPFNCNCHLAWFAECVRKKSLNGGA--ARCGAPSKVRDVQIKDLPHSEFKCSS----- 703

Qy    315 PPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVM 350
                :||   :  |  | : |        |||:
Db    704 -----ENSEGCLGDGYCPPSCTCT--------GTVV 726
```

```
RESULT  14
US-09-182-024A-5
; Sequence 5, Application US/09182024A
; Patent No. 6342370
; GENERAL INFORMATION:
;  APPLICANT:  Connolly, Timothy
;  APPLICANT:  Rajput, Bhanu
;  TITLE OF INVENTION:  Human Slit Polypeptide and Polynucleotides Encoding
;  TITLE OF INVENTION:  Same
;  FILE REFERENCE:  640100-271
;  CURRENT APPLICATION NUMBER: US/09/182,024A
;  CURRENT FILING DATE:  1998-10-29
;  PRIOR APPLICATION NUMBER:  60/063,946
;  PRIOR FILING DATE:  1997-10-31
;  PRIOR APPLICATION NUMBER:  60/096,420
;  PRIOR FILING DATE:  1998-08-13
;  NUMBER OF SEQ ID NOS:  5
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 5
;   LENGTH: 1480
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-182-024A-5


  Query Match             13.7%;  Score 347;  DB 4;  Length 1480;
  Best Local Similarity   23.9%;  Pred. No. 3.3e-20;
  Matches  109;  Conservative   65;  Mismatches  152;  Indels  130;  Gaps    9;

Qy      2 TCPSVCSCSNQFSKVICVRKNLREVPDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEIL 61
          :||  | |::    | |  |:|  ||  :  :|  : | :| |  :   ||   | |  :
Db    294 SCPHPCRCADGI--VDCREKSLTSVPVTLPDDTTDVLLEQNFITELPPKSFSSFRRLRRI 351

Qy     62 QLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNNPIESIPS 121
          || |:|  |   | :||  | || |: |::  :|:| |  |  |: | |  | | |
Db    352 DLSNNNISRIAHDALSGLKQLTTLVLYGNKIKDLPSGVFKGLGSLRLLLLNANEISCIRK 411

Qy    122 YAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAM------CNLR--------- 166
          ||  : ||  | | :   :  :: | |: : ::: ::||       ||||
Db    412 DAFRDLHSLSLLSLYD-NNIQSLANGTFDAMKSMKTVHLAKNPFICDCNLRWLADYLHKN 470

Qy    167 ---------------------------------------------------------- 166

Db    471 PIETSGARCESPKRMHRRRIESLREEKFKCSWGELRMKLSGECRMDSDCPAMCHCEGTTV 530

Qy    167 --------EIPNLTPL--------------IKLD----------ELDLSGNHLSAIRPGS 194
                  |||   ||             |  |          :|:|  | |: | | :
Db    531 DCTGRRLKEIPRDIPLHTTELLLNDNELGRISSDGLFGRLPHLVKLELKRNQLTGIEPNA 590

Qy    195 FQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLTLLPHDLFTPLHHLERIHL 254
          |:|  |:|:| : :::|: |    |  |  |  :||  | :: :    |  |: |  ::|
Db    591 FEGASHIQELQLGENKIKEISNKMFLGLHQLKTLNLYDNQISCVMPGSFEHLNSLTSLNL 650

Qy    255 HHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFTCYAPVIVE 314
            ||:|||| : | :  ::  : :  |  |||  |  ::   | :|  : | | :
Db    651 ASNPFNCNCHLAWFAECVRKKSLNGGA--ARCGAPSKVRDVQIKDLPHSEFKCSS----- 703

Qy    315 PPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVM 350
              :||   :  |  | : |        |||:
Db    704 -----ENSEGCLGDGYCPPSCTCT--------GTVV 726


RESULT  15
US-08-592-500-2
; Sequence 2, Application US/08592500
; Patent No. 6005089
;  GENERAL INFORMATION:
;    APPLICANT:  Lanza, Francois
;    APPLICANT:  Phillips, David R.
;    APPLICANT:  Cazenave, Jean-Pierre
;    TITLE OF INVENTION:  Platelet Glycoprotein V Gene and Uses
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  379 Lytton Avenue
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/592,500
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/089,455
;      FILING DATE:  09-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dow, Karen B.
;      REGISTRATION NUMBER:  29,684
;      REFERENCE/DOCKET NUMBER:  12418-28
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 326-2400
;      TELEFAX:  (415) 326-2422
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  560 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-592-500-2


  Query Match             13.1%;  Score 329.5;  DB 3;  Length 560;
  Best Local Similarity   29.7%;  Pred. No. 2.4e-19;
  Matches   83;  Conservative   43;  Mismatches  104;  Indels   49;  Gaps    3;

Qy     22 NLREVPDG-----------------ISTNTRLLN---------LHENQIQIIKVNSFKH 54
          ||  :| |                  :| :: |||          | | |: |   :|
Db    181 NLTHLPKGLLGAQAKLERLLLHSNRLVSLDSGLLNSLGALTELQFHRNHIRSIAPGAFDR 240

Qy     55 LRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIPNGAFVYLSKLKELWLRNN 114
          | :|  | |||||:  :    |    ||  | ||:| |  :|   |  :  |:||||
Db    241 LPNLSSLTLSRNHLAFLPSALFLHSHNLTLLTLFENPLAELPGVLFGEMGGLQELWLNRT 300

Qy    115 PIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLREIPNLTPL 174
          : ::|: ||  :  || | :    ||| : :|||:||  |:
Db    301 QLRTLPAAAFRNLSRLRYLGVTLSPRLSALPQGAFQGLGELQV----------------- 343

Qy    175 IKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNN 234
              | |  | |:|:  |  :||  |::: : ::::: : |  | || ||  : | ||
Db    344 -----LALHSNGLTALPDGLLRGLGKLRQVSLRRNRLRALPRALFRNLSSLESVQLDHNQ 398

Qy    235 LTLLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIK 273
          |  || |:|  |  |  : | || | |:| :     |::
Db    399 LETLPGDVFGALPRLTEVLLGHNSWRCDCGLGPFLGWLR 437


Search completed: August 26, 2002, 15:46:38
Job time: 170 sec
```